万卷方法

HANDBOOK OF
RESEARCH DESIGN
&
SOCIAL
MEASUREMENT

研究设计与社会测量导引

第6版

德尔伯特·C·米勒
内尔·J·萨尔金德
著

风笑天 等
译

重庆大学出版社

封面设计/黄河封面工作室 huang_he@21cn.com

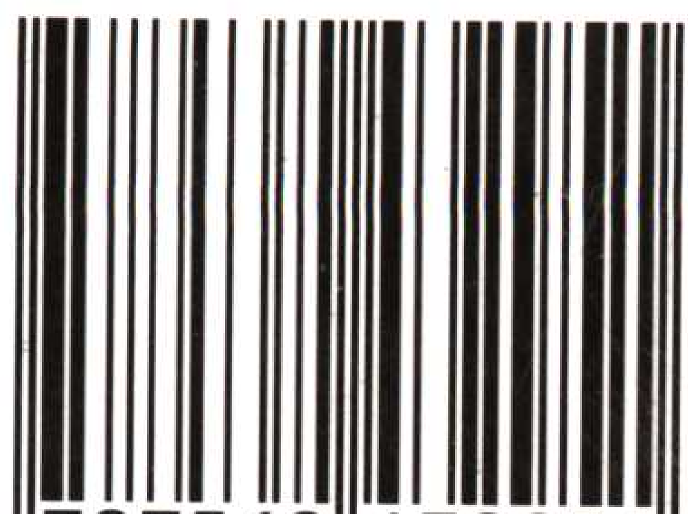
这本方便日常研究查阅的工具书

- 系统介绍了社会科学应用研究、基础研究、评价研究的基本认识及研究设计的程序和方法（第1～3章），定性研究方法的五个重要领域——叙事研究、现象学、扎根理论、民族志、个案研究（第4章），收集、统计和分析数据的方法、工具和模型（第5～6章），以及量表和指标的选择与运用（第7章）。
- 提供了进一步学习和阅读的资料，如书籍、杂志、论文、网站、学术团体、出版社等。
- 全文或部分选载了关于研究方法与研究设计方面的重要文献或工具。

向您推荐

- 《美国社会学学会伦理准则》及其应用实例《向堪萨斯大学人类咨询委员会呈交议案的说明书》（pp. 102～131）。
- Elliot W. Eisner教授在定性研究特别兴趣小组十周年庆祝会上的一个长篇讲话：《定性研究方法论的新领域》（pp. 151～159）。
- “父亲的职业获得、教育获得与儿子的教育、首次职业、最近职业间的路径关系”这一路径分析方法应用的著名案例（pp. 256～261）。
- M. W. Lipsey & D. B. Wilson的作品选载：《应用元分析》（pp. 273～279）。
- 大量可以借鉴的量表和指数：Hemphill群体纬度指数量表及其使用（pp. 305～313）、近150个社会指标的定义（pp. 320～327）、组织内与工作相关紧张/焦虑指数量表及其使用(pp. 354～356)、社区态度量表及其使用(pp. 357～360)、闲暇活动调查表（pp. 365～366）、管理行为描述量表（pp. 371～373）、社会生活量表（pp. 375～379）等，以及《心理测量年鉴》（MMY）多年来发表的2000个左右的心理测量量表的查找索引（pp. 391～448）。

ISBN 7-5624-3295-3



ISBN 7-5624-3295-3
定价 68.00

万卷方法

HANDBOOK OF RESEARCH DESIGN & SOCIAL MEASUREMENT

研究设计与社会测量导引

第6版

德尔伯特·C·米勒
内尔·J·萨尔金德
著

风笑天 邹宇春 邓希泉 史伟伟
吴元清 李良进 蔡扬眉 张静波
杜荟苑 童宗斌 江 臻
译

重庆大学出版社

图书在版编目(CIP)数据

研究设计与社会测量导引/(美)米勒(Miller, D. C.),(美)萨尔金德(Salkind, N. J.)著;风笑天等译.—重庆:重庆大学出版社,2004.12

(万卷方法)

书名原文:Handbook of Research Design & Social Measurement

ISBN 7-5624-3295-3

Ⅰ.研… Ⅱ.①米…②萨…③风… Ⅲ.社会测量—研究 Ⅳ.C915

中国版本图书馆CIP数据核字(2004)第131075号

研究设计与社会测量导引(第6版)

德尔伯特·C·米勒 内尔·J·萨尔金德 著

风笑天等 译

责任编辑:雷少波 版式设计:周 晓

责任校对:邹 忌 责任印制:张 策

*

重庆大学出版社出版发行

出版人:张鸽盛

社址:重庆市沙坪坝正街174号重庆大学(A区)内

邮编:400030

电话:(023) 65102378 65105781

传真:(023) 65103686 65105565

网址:http://www.cqup.com.cn

邮箱:fxk@cqup.com.cn(市场营销部)

全国新华书店经销

重庆科情印务有限公司印刷

*

开本:787×1092 1/16 印张:29 字数:635千 插页:2

2004年12月第1版 2006年2月第2次印刷

ISBN 7-5624-3295-3 定价:68.00元

本书如有印刷、装订等质量问题,本社负责调换

作译者简介

德尔伯特·C·米勒(Delbert C. Miller) 印第安那大学社会学与商业管理方面的教授。米勒教授在专业学术期刊上发表了大量的研究论文,还与 William H. Form 教授合作了 Industrial Sociology(本书共出版了 3 个版本)及 Industrial, Labor, and Community 两部书。他的两部个人专著 International Community Power Structure 与 Leadership and Power in Bos - Wash Megalopolis,描述了他直接指导的大规模团队研究项目。另外,作为一名监督培训专家及仲裁员,米勒教授为美国仲裁协会及印第安那州服务多年。

内尔·J·萨尔金德(Neil J. Salkind) 堪萨斯大学心理与教育研究方面的教授,过去的 29 年他一直在这所大学教授儿童发展与研究方法课程。萨尔金德教授主要的研究兴趣在政策分析及影响青少年发展的政策方面。萨尔金德教授从马里兰大学获得哲学博士学位,是美国心理学协会及儿童发展研究会的会员,他从事儿童发展研究摘要及文献目录的编辑工作达 13 年之久。

风笑天 南京大学社会学系教授,博士生导师。1990 年 6 月毕业于北京大学社会学系,获法学博士学位。现兼任教育部社会学学科教学指导委员会委员,中国社会学会常务理事,中国社会学会社会研究方法专业委员会理事、副秘书长,中国人民公安大学荣誉教授,华中师范大学、华中农业大学、广西师范大学、浙江师范大学兼职教授。出版有《社会学研究方法》、《现代社会调查方法》、《独生子女:他们的家庭、教育和未来》、《社会调查中的问卷设计》等著作,在《中国社会科学》、《社会学研究》等学术刊物上发表论文近百篇。主要研究领域包括:社会学研究方法、青年社会学、家庭社会学、中国独生子女问题等。

为方法理性鼓与呼*

——万卷方法策划报告暨出版说明

究竟是中国内地学界对于研究方法的漠视,导致研究方法出版物的匮乏?
还是研究方法出版物的匮乏,导致学界没有对方法问题予以应有的重视?
这是一个鸡生蛋蛋生鸡的问题。
作为图书出版的编辑人和策划人,对之多论无益。
但,作为图书出版的编辑人和策划人,我们却非常乐意——为方法理性鼓与呼!

我们乐于,也更善于从出版物的数量和质量的比较中,来了解和表述某一类知识的生产和传播情况,以此作为我们图书出版策划的基础。同样,在万卷方法的策划之前,我们便对美国、中国台湾及中国内地三者,1999—2002年间关于社会科学研究方法的图书出版情况做了一个简单的比较:

美国在社会科学研究方法方面出版物的数量之多,至少足以让我们惊讶!由于不便统计,在此,仅就美国的SAGE出版公司在社会科学研究方法方面的图书出版情况,做一个简单介绍,以窥其一斑。SAGE出版公司4年中出版的社会科学研究方法方面的书籍(包括再版书)便有两百余种,其中,既有一两百万字的大部头工具书,如Handbook of Qualitative Research,Handbook of Research Design & Measurement;也有10万字左右的口袋书,如仅一个"应用社会科学研究方法系列丛书"(Applied Social Research Methods Series)便有49个品种,其中每本书对于案例研究方法、调查研究方法、网络调查方法等做了分门别类的介绍。其覆盖面之广,研究之深入、具体、操作性强是我们所难以想象的。如果说,我们与美国的差距尚可理解和接受的话,那么,同一时期,中国台湾在这一方面的研究和传播情况也足以令我们汗颜。4年间中国台湾出版了大量深入研究社会科学研究方法方面的书,其中不但有本土作者的原创作品,也有从英美等西方国家引进版权的相对比较成熟的社会科学研究方法方面的作品。更为可贵的是,其中有几家出版公司已经注意从社会科学研究方法的体系着眼来组织自己的策划和出版,在这方面的出版有了一定的规模,内容的组织也显得比较成熟:如韦

---

* 本文原载《中华读书报》2004年3月31日,第15版。在此,根据需要作者对原文做了少量改动。

伯文化事业出版社的“研究方法系列丛书”、弘智文化事业公司的“应用性社会科学调查研究方法系列丛书”等都是发展比较成熟、规模比较大的专门传播社会科学研究方法的系列丛书。

而中国内地方面，同期虽然也出版了一批研究方法的书籍，但无论是规模还是研究内容的深入丰富程度都无法与美国及中国台湾两地相提并论。从规模看，这期间中国内地出版的社会科学研究方法类的著作也就四五十本，难望美国同行之项背，仅与中国台湾同期出版规模大致相当。从此类出版物内容的深入丰富程度来看，大陆出版的社会科学研究方法类出版物主要集中在各学科内部，如教育研究方法、心理学研究方法、社会学研究方法、经济学研究方法、体育科学研究方法等，每本书都面面俱到谈调查、谈案例、谈访谈、谈田野、谈观察等，而很少有对某一种方法进行深入研究的图书，更没有像美国或中国台湾那样形成比较系统的研究社会科学研究方法的系列丛书。在这一领域，这样的图书结构对于应付大学本科生在研究方法方面的入门性需求(即作为教材)还行，但对于迅速培养一大批拥有科学、理性的研究头脑的学者，特别是对于青年学者，如硕士、博士研究生的成长则是远远不够的。其致命的弱点表现在三个方面：其一，一本书铺得太开而无法细化深入，以至于绝大多数学生虽然上了方法的课，却除了记住了几个名词和对一些方法的理论及应用略知一二之外，至于什么情况下选择什么方法最恰当、具体怎么操作、这种方法可能会有哪方面的不足需要加以处理等研究细节问题仍然处于无理性状态。其二，既然我们的方法建设囿于学科内部，而不能对方法进行纵深的开发，那么，反正学科就那么几个，于是方法书大多是低水平重复。比如教育研究方法的书，现在全国使用做教材的至少也有十多本吧，但你若有心思看的话，一本足矣！其三，出版界对研究方法图书这种淡淡的一笔带过的做法，不仅很难形成一种风气，从意识深处提升起大家对于研究方法的关注，而且更加剧了整个学界对于研究方法的漠视，许多学者只有在著书立说之时才想到似乎从“作品结构的完整性”上看应该谈及方法问题。

社会科学界近年流行两个词语：一曰反思，二曰接轨。所谓反思也即是对于学科的理论基础、学科的结构、学科的历史与未来等问题的全面梳理；所谓接轨也即是希望中国的社会科学研究能够融入世界社科研究的大潮中，与国际上的社会科学研究实现彻底的、清晰的对话。在此，我们以为，无论是反思还是接轨，中国的社会科学界和传播界都必须投入一些精力来关注、研究、教授和传播社会科学研究方法。惟如此，才能在中国形成完善、科学的社会科学研究方法体系和学科群，也才能使对研究方法关注与理性应用在中国社会科学界深入人心、蔚然成风；惟如此，也才能为中国社会科学研究与国际接轨搭建一个平台。

以上种种，促使我们策划出版了万卷方法系列丛书，献给忠实于人文社会科学研究的人们！

雷少波　崔祝
2004 年 6 月于重庆大学出版社

前　言

第6版是这本书历史上修改得最多的一版。我们一如既往地强调社会科学家必须掌握的,同时也是使他们能够进行高质量经验研究的各种工具。当然,我们也认识到,社会科学家必须能够适应迅速变化的、充满大量信息的世界。

与前几版一样,这本书的第6版为社会研究的各个阶段都精心设计了资料读物。它力图为社会与行为科学中的研究以及新闻、教育、社会工作、商业等应用性专业提供指导。而酒店管理、健康服务、公共卫生、城市管理以及法律实施等部门的研究者也将会发现本书同样适用于他们的问题。

到这一版为止,这本书荣幸地为社会与行为科学研究者服务了三十多年。它的用途主要体现在以下三个方面:

1)对于社会及行为科学的研究者来说,它是一本独立的参考书。作为一本参考书,这本书中包括了许多有用知识的摘要,它们是社会及行为科学研究者十分必要的研究指南。本书在内容上涵盖了对各种期刊、有用的资料读物、参考书目、计算机软件、许多最重要的量表、基金和出版资源以及其他许多材料的搜寻,目的是为了使书中有足够的信息,以便让有经验的研究者也能够发现在他的研究生训练阶段所没有学过的材料的摘要,以及可以被其他任何研究者使用的信息。

这一修订版中也认识到定性研究方法的价值,并为此而增加了整整一章的篇幅。

2)对于那些希望给他们的学生提供有关社会科学研究及其构成的教师来说,本书可以作为社会科学研究方法教科书的补充。

3)最后,它可以作为一本教科书,为学生的研究训练提供一步一步的指南。它从每一个基础研究者——学生或职业研究者——的起点开始,寻找一个有创造性的观点,一个中层理论(middle-range theory),以及初始的假设;然后进行研究设计,撰写课题申请报告,收集和分析资料;最后撰写和出版研究报告以及如何利用研究技术。

对于应用研究者和评价研究者来说,本书对他们在研究工作中所面临的实践问题也进行了概括的介绍。对基础研究、应用研究、评价研究的活动和目标也都进行了描述。本书还描述了通过各种理论和方法论工具来掌握这些问题的方式。

在力图保持前期版本的标准的同时,这一修订版包含了大量的新材料。主要有:

- 对定性研究方法的综合介绍。
- 新的、更及时的评论和解说取代了过期的或不再相关的摘录。
- 关于重要的社会科学变量处理的数以千计的新文献以及有关统计分析、计算机应用及其他一些特定主题的文献。
- 有关计算机应用和网络研究技术的最新信息,包括有关互联网及在现代科学研究者中运用的初始的和中间的材料。
- 对大范围的学术机构的活动和程序的评论。
- 有关社会与行为科学研究中伦理事务的讨论和指南。
- 为包括国际范围中的情况而对社会指标的扩充。
- 本书历史上第一次出现的一个综合的指数。

这一版是 Delbert C. Miller 教授与 Neil J. Salkind 教授之间协作的开端。令这么多年来一直与这本书为伴的每一个人悲痛的是,当修订工作刚刚开始的时候,Miller 教授却不幸去世。他关于这本书应该成为一本对大量初学者和高级研究者来说什么样的资源的眼光,给这一修订版以深深的启示。将这本书第 5 版中 Miller 教授的致谢辞在此重述,既是必要的,也是恰当的:

> 这些年来,在我身后有三位社会学教授用同一种思想给我以启示:社会的资料就是自然的资料,对它们的研究能够而且应该遵循同样的科学方法。我作为一名物理学和数学教师在高中教了两年书之后,又成为了一名社会科学的大学生。我回到了大学,学习社会研究的知识。现在,我必须向我在迈阿密大学的硕士导师 Read Bain,在明尼苏达大学的博士导师 F. Stuart Chapin 以及他的学生,后来成为我的老师和华盛顿大学社会学系系主任的 George A. Lundberg 致敬。他们都具有一种把社会计量尺度作为社会测量基本元素的眼光。至于本书,则是带着这种眼光产生于宾州大学开设的一门社会测量课程的教学实践。它经过了我在印第安纳大学 31 年的培育。1951 年至 1987 年发表在美国社会学协会的官方刊物《美国社会学评论》(American Sociological Review)上的每一篇研究论文,都经过了所用尺度的考察,并按照类型、频数和趋势进行了分类。在这本书的这一版和前一版中,可以看到总结性的注释。

对于 Miller 博士,我深怀感激!是他给了我参与这一修订版的机会。我希望这本书的价值能够体现出他过去所作出的贡献。

致　谢

如果一个作者认为，一本书从最初的想法到摆在书架上都是作者的功劳，那么他是无知的，特别是对于这么大量的修订来说，更是如此。C. Deborah Laughton，本书的编辑，绝对是一个随时准备以任何必需的和可能的方式给予作者以支持和帮助，直到书稿最终完成的人。如果这本书有任何成功之处，那么，在很大程度上都归功于她的专业素养、她的能力以及她对作者的真诚的关心。在这个编辑更换公司和工作就像季节变换一样频繁的年代，我们还是可以肯定地知道，在出版界，仍然有人信仰连续性和完整性。对她，我充满了感激！

我还要特别感谢 A. J. Sobczak 和 Diana Axelsen。A. J. 在名义上只是一个文字编辑，但在实践中却远不止于此。他核对事实，寻找参考文献，建议结构方面的调整，并确保每一部分中的数字都准确无误，每一处修改都确定无疑。他是大量建议和持续鼓励的来源。Diana 管理着整个的生产过程，转换着书稿的流程，将它变成一部吸引人的和有用的著作。她的耐心贯穿着整个生产过程，她留心每一处细节，温和的鼓励洒满书页。Janelle LeMasster 设计了本书的内部装帧，并极其沉着地处理了复杂的排版工作。

最后，我们要感谢所有的评论者：北卡州立大学的 Michael L. Vasu，爱荷华大学的 Dan Berkowitz，弗尔曼大学的 Dan Cover，布莱汉姆·扬大学的 Barry L. Johnson，罗德岛学院的 Jianhong Liu，圣母玛丽亚大学的 David M. Klein，夏威夷大学 Manoa 校区的 Ronald H. Heck，南加州大学的 David Lopez-Lee，以及卡尔加里大学的 Marja J. Verhoef。

Neil J.Salkind
堪萨斯大学
劳伦斯，堪萨斯 66044

目 录

1

INTRODUCTION: UNDERSTANDING BASIC, APPLIED, AND EVALUATION RESEARCH

导论:理解基础研究、应用研究和评价研究

第 1 部分定义了基础研究、应用研究和评价研究的特征,包括对三种研究类型的一个简要的讨论。同时还从机遇与局限的角度对这些方式中每一种进行了描述。

1.1 基础研究、应用研究与评价研究的特征

对行为或组织问题感兴趣的研究者可以将他们的需求集中在三个不同的方向上，这三个方向通常被称作基础研究、应用研究和评价研究。尽管它们有时根据具体运用背景的不同，而被用于其他的名称之下。例如，基础研究经常被称作纯研究或者实验研究；应用研究经常出现在政策研究和行动研究的术语中；而评价研究有时则被归为评估研究或估算研究，甚至被称作社会会计学。

尽管有不同的术语来描述这三者，但我们在本书中，还是采用基础研究、应用研究和评价研究的名称。

通过提供三者的基本特征，表 1.1 从所探讨问题的性质、研究的目标、潜在的或指导的理论，以及研究中适用的技术四个方面展示了基础研究、应用研究和评价研究三者之间的某些差别。

表 1.1 研究设计的取向

特征项	基础研究	应用研究	评价研究
问题的性质	基本的科学考察，寻求有关社会现象的新的知识，希望建立用以解释社会现象的一般性原理和理论	应用研究希望理解基础研究的结果，如何能够帮助减轻社会问题的压力和为决策者提供有充分根据指导的补救措施	评价研究评定对社会问题进行处理的结果或者成功实践的结果
研究的目标	产生新的知识，包括发现变量之间关系的性质	在对那些希望排除或减轻社会问题的决策者有用的应用背景中探索基础知识的价值	提供一种对应用于社会问题的措施的精确评估
潜在的理论	选择一种理论用以指导假设检验	选择一种理论、方针，或者是直觉的预感来探讨社会系统的动力	选择一种理论以适合所评价的问题
适用的技术	理论表述、用实验的和准实验的方法检验假设、抽样、资料收集技术、数据的统计处理、证实或证伪假设	非常类似于基础研究，只是处在一个研究的应用可以立竿见影的背景中	运用适合问题的常规技术

仔细观察表 1.1 中的内容，不难发现某些连接三种研究类型的线索。例如，基础研究所寻求的是有关社会现象基本问题的答案中的新知识，应用研究则希望表明这种知识如何能够被用于处理某种紧迫的问题，而评价研究则包括在有关一个项目的价值或优点的判断之中。

每一种研究类型都可以独立进行，但任何一种研究类型的价值都必须根据其他两

种研究类型的结果来判断。越来越多的研究者发现他们往往处于不同的阵营中。尽管不同研究类型在研究目标上的分歧为研究者提供了不同的研究取向,同时,研究者也强烈地感觉到其研究类型的价值,但是,要以最好的方式回答大多数研究问题,所有的社会与行为科学家都必须理解每一种研究类型。

研究者对基础知识、理论、设计、方法论技术以及一般的应用掌握得越好,准确回答特定研究问题的机会就越大。无论是发现问题、确定问题、发现可能的设计、资料的收集和处理,还是假设的最终检验,都是如此。

尽管这本书基本上是为着理解基础研究的,但是,也将对应用研究和评价研究的重要领域具有参考作用(特别是在第2部分和第3部分中)。在这些重要领域中活跃着社会实践者,并且他们的思想每天都在接受检验。Rossi 和 Wright(1985)对不同研究领域之间的这种互惠关系作出了如下说明:

> 当应用研究者试图解决更为复杂的应用研究任务时,我们同样预期基础研究将继续从应用研究者所作出的大量的技术进步中获得好处。毫无疑问,许多技术和大量的知识会以同样的方式从基础研究流向应用研究者。一个没有应用领域的学科就失去了理论和方法的某种丰富性;而一个失去了与基础学科之间联系的应用领域,也存在着这样一种风险,即对决策者关于社会问题及最可行的解决办法的阐明的关注有失偏颇和过于狭窄。(p. 76)

参考文献

Rossi, Peter H., & Wright, James D. (1985). Evaluation research: An assessment. In Linda H. Aiken & Barbara H. Kehrer (Eds.), *Evaluation Studies Review Annual* (Vol. 10). Beverly Hills, CA: Sage.

基础研究的取向和责任 1.2

从事基础研究的大多数研究者都将他们的目标放到发展知识或者理解基本的社会现象上。任何应时的和明显的效用对于科学的基本目标来说都是第二位的。科学要描述世界的本来面目,而不在意它应该如何被改变。尽管大部分人都同意,所有的知识,无论它是否验证了一种假设,最终都将是有用的。然而,认识到什么是真实的,什么不是真实的,却永远是有用的。Hans Selye,一位发展出现代压力概念并且研究了压力对人体影响的内分泌学家,他的论述与这一论点一致。他始终认为,许多伟大的科学家做出无数有价值的贡献,主要不是被他们的好奇心或者为人类服务的强烈愿望所激发,而是被一种不愿接受失败的情感所激发。

获得这种基础知识的方式体现了无数的挑战。在现实世界中,大量的社会变量相互之间是高度相关的,同时,明确的因果关系往往难以建立。对于同一种现象,充满了各种不同的理论和解释,研究者中存在的各种偏见和在意识形态上的差异,也是导引

我们寻求答案的各种方法论视角的一部分。

那么,为什么社会研究者要使他们自己去同这样一种不确定的、狡猾的、令人糊涂的,同时还是令人恼怒的概念和思想打交道呢?许多人欢迎在一个新的科学领域工作所具有的挑战,原因是它的发展还不足一个世纪;另一些人信奉社会科学应研究他们认为的我们这个时代最重要的问题的取向,同时他们还把社会科学看作是解决问题的希望所在;还有一些研究者则敬重科学方法的诚实性,以及那些需要智力训练与诚实性的数学的和概念的工具。

有责任感的社会研究者不断地追寻知识,而无论结果如何。最经常的提问是对关系的考察,并试图使其适合这样一种一般性的命题:"在一定的条件下,如果X发生变化,那么,Y也会发生变化。"作为好的科学家,这些研究者努力复制和确认结果,随着预测能力的每一点增加,知识的价值、学科的价值也在增加,结果与事物的"真实的"状况也越接近。当一个决策者需要从社会科学中得到基本资料,并且这些资料都是不可缺少的,社会科学就会向前发展。同时,理论、资料与研究的相互交织,会增强理论,社会研究就会成为社会科学。

1.3 应用研究的取向和责任

应用研究者寻求能够用于解决紧迫的社会和组织问题的知识。这种知识通常是基础研究的一种扩充,它应该是有效的、描述的、可告知如何进行改变的。

政策分析或者行动研究者经常认为,我们所具有的关于社会科学现象的知识还没有通过使用合适的方法论工具来产生,因此,我们对社会组织结构下的行为动力的准确解释是有一定局限的。不过,大多数应用研究者都愿意接受"应用的思想形式"(applied mind-set),即努力将研究的资料用于应用性的背景中。

1.3.1 应用研究的类型学

Rubin(1983)发展出一套做应用性社会研究的社会科学家的类型学,见表1.2。

表1.2 应用性社会研究的社会科学家表现

	聚焦于政策应用	聚焦于资料收集和资料解释
关注社会过程	政策分析者	评价研究者
关注社会结果	社会监控者	资料分析者

- 政策分析者研究社会过程,描述对于解决现存问题来说存在哪些政策选择。
- 评价研究者研究社会过程,以决定一个项目或一项工程是否达到了它所希望达到的目标。
- 社会监控者考察结果资料,以便发现各种需要某些组织或者政府行动的模式。
- 资料分析者运用和改进方法论的工具来解释结果资料,经常运用高级统计程序。

根据人们寻求帮助的愿望以及所考虑的问题的需要,一个应用研究者可能会被指派担当上述一种或者全部角色。此外,所有的应用研究者都要为理解研究方法论承担职责,同时把决定收集资料的费用与资料的价值作为决策过程的一部分。

大学提供某些应用研究方面的课程,但是最有可能学习应用研究技术的地方,是在社会科学研究所的研究活动中。在那里,学习是通过在作为实习医生、学徒以及刚刚参加工作的职员的工作中进行的。专业学院比如商学院、社会工作学院以及教育学院都容易接纳应用研究,认为其"在理论和实践中都有用"。在那些学院的"实践性"研究中,可以看到许多特殊的教授。

1.3.2 作为入侵者的应用研究者

应用研究者经常被看作(无论好坏)是在一个通常相对紧密的社会系统中变革的代理人。每一个人似乎都对应用研究者在干什么,他或者她在寻求什么,以及为什么他或者她"潜伏在附近"(lurks around)等等充满了好奇,并对之提出了很多的问题。另一些人则开始问道:"你们希望发现什么?""你们打算干什么?""你打算要我们躺到床上吗,医生?""经理试图要我们增加产量吗?"

管理者可能拒绝合作,或者带着愠怒和不关心进行合作。如果最高层的管理者支持研究的态度不真诚,那将会有无穷尽的困难。任何身处敌意环境中的研究者都会发现应付起来非常困难。这就是为什么我们鼓励我们的学者在应用研究过程中要积极参与的原因,这种积极参与有助于建立起对应用研究的理解和宽容。

研究者存在着这样的想法,"我想要帮助改善组织中个体的福利,为什么雇员和经理如此戒备?他们对我的研究为什么如此怀疑?"这些问题应该集中在应用研究者的目标上,同时应该对任何可意识到的侵扰的目的进行说明。研究者应该准备回答这些问题,同时为研究提供恰当的理由以便获得雇员和经理的充分支持。

参考文献

Rubin, Herbert J. (1983). *Applied social research.* Columbus, OH: Charles E. Merrill.

1.3.3 应用研究者未来训练的需要

美国社会学协会(ASA)充分意识到,社会学创造与生产目标相联系的合适的知识的能力,以及提供对评价和决策过程有价值的知识的能力,应该受到重新评估。美国社会学协会定期主办有关应用社会学方向的研讨会。若需要了解相关信息或者取得参与资格,可在美国社会学协会的网页(www.asanet.org/)上查询特别项目。

目前方法论的训练强调计算机分析技术和掌握适当的分析资料的软件,以便充分地回答最初形成的研究问题。评价研究和决策研究具有高度的优先性。

需要记住的是,应用研究训练和基础研究训练并不是完全相互排斥的,特别是对社会科学家的训练,它们是高度互补的。受过良好训练的社会科学研究者可以有效地参与应用研究和基础研究。事实上,最好的应用研究者具有坚实的基础研究训练。基

本的差别在于研究所提出的问题以及研究的目标——而不一定是探究问题时所选择的方法。

大量的联邦基金用于应用行为科学家和应用研究者的训练,近年来,由于联邦政策的改变,基础研究(传统上一直资金不足)和应用研究的基金都在增加。

1.3.4 应用研究组织的不断出现

社会学的实践者、社会学的实践组织以及应用研究自身数量的增长,证实了对应用研究和社会学实践的兴趣的不断增长。社会学实践协会或简称 SPA(www.socpractice.org/)致力于社会学的应用,已有二十多年的历史。社会实践协会是一个实践者和学者的社区,在那里,成员们把他们的干预技巧与研究、技术以及批评性分析相结合。社会实践协会的使命是:

- 通过开展学者间的和教育的活动,促进社会学知识对个人和对社会变迁的应用。
- 开发社会学实践者就业和发挥作用的机会。
- 为社会学实践者、相关的专业人员以及对此有兴趣的学者和学生提供一个共同的交流场所。
- 为进一步的社会学实践增进训练和教育的机会。
- 提高理论、研究和方法水平,以改进社会学方法对变迁的作用。

作为应用社会学中一个兴趣在持续增长的方向,社会学实践协会目前出版了《社会学的实践:临床与应用社会学杂志》(*The Sociological Practice: A Journal of Clinical and Applied Sociology*)。这本杂志包括有关社会学临床运用的论文、评论、实践以及研究报告。第1期(1999年3月)中包含下列文章:

"Introductory Statement: Philosophy and Future Direction," by John G. Bruhn

"Tomorrow's Most Challenging Practice Issues: 21st Century Surprises," by Arthur B. Shostak

"Installing Circuit Breakers: Mechanisms to Avoid Sexual Harassment in the Workplace," by John Marker

"Gambling, Drugs, and Sex: New Drug Trends and Addictions in Honolulu, Hawaii, 1998," by Gordon James Knowles

"What Sociologists Do and Where They Do It-The NSF Survey on Sociologists' Work Activities and Workplaces," by Robert J. Dotzler and Ross Koppel

"Book Review: *Developmental Theories of Crime and Delinquency*," edited by Terence P. Thornberry, reviewed by Blake Nelson

其他集中关注应用社会学的组织包括应用与临床社会学委员会(www.sociologycommission.org),其目的是"在应用和临床领域发展、促进和支持高质量的社会学教育与实践",以及应用社会学学会或简称 SAS(www.appliedsoc.org/),这是"一个在广泛的背景中应用社会学知识的专业人员的国际性组织"。应用社会学学会的目的是:

- 为社会学家以及对应用社会学知识感兴趣的人提供一个论坛。

- 增强对社会学知识与社会学实践之间内在联系的理解。
- 提升应用社会学研究和训练的效果。

评价研究的取向和责任 1.4

评价包括对项目、政策、人事、产品以及组织的长处和不足进行评估,以努力改进其效果。在一项政策或项目需要进行评估时,就会邀请评价研究者参与。

从事任何评价训练的一个不可缺少的部分是对所研究的活动是否有价值作出判断。Edward Suchman(1967)确定了一组可以帮助做出这种判断的标准:

- 努力或活动的水平。
- 表现或完成的水平。
- 项目的适当性或效果。
- 相对于投入的产出或效率。
- 对项目有效性的条件的说明。

根据 Suchman 的观点,前两条标准是评价性的,并且关注的是活动与结果之间关系的确定;第三与第四条标准实际上是管理性的,它涉及相对于结果来说,所作出的努力大小与花费之间的判断;最后一条标准注重增加知识和理解,而不考虑结果。

当前的政治气候以及近来对责任重要性的关注,指出了应用社会学家对资源的需求。评价者可以向美国评价协会(www. eval. org/)求助。这个协会是一个国际性的评价者的专业协会,它主要关注于"项目评价、人事评价、技术评价以及许多其他形式的评价的应用和探讨"。它的使命是:

- 改进评价实践与方法。
- 增加评价的利用。
- 推动评价作为一种职业。
- 支持评价在产生有关有效的人类行动的理论和知识中的贡献。

基础研究、应用研究与评价研究方面的读物推荐 1.5

通读本书,你将会发现本书提供了大量的对于不同领域的社会科学家来说,从事相关研究的重要的研究方法文献。下面是有关基础研究、应用研究及评价研究方面的重要而有趣的文献。

1.5.1 基础研究

Bailey, Kenneth D. (1997). *Research methods for creating a just world: Overcoming barri-*

ers to inclusion in social research. Los Angeles: University of California; Society for the Study of Social Problems.

Blessing, Lusienne T. M. , & Chakrabarti, Amaresh. (1999). *A design research methodology.* Godalming, Surrey, UK: Springer Verlag London.

Dyer, James A. (1996). *The art and science of survey research.* In Timothy J. Flanagan & Dennis R. Longmire (Eds.), *Americans view crime and justice: A national public opinion survey.* Thousand Oaks, CA: Sage.

Lacy, Michael G. (1997). Efficiently studying rare events: Case-control methods for sociologists. *Sociological Perspectives*, 40, 129-154.

Tabachnick, Barbara G. , & Fidell, Linda S. (2000). *Computer-assisted research design and analysis.* Needham Heights, MA: Allyn and Bacon.

1.5.2 应用研究

Bechofer, Frank, & Paterson, Lindsay. (2000). *Principles of research design in social sciences.* New York: Routledge.

Grunow, Dieter. (1995). The research design in organization studies: Problems and prospects. *Organization Science*, 6, 93-103.

Kenny, David A. (1996). The design and analysis of social-interaction research. *Annual Review of Psychology*, 47, 59-86.

Narver, Betty Jane. (1999). Applied research: Entrepreneurial style. *Metropolitan Universities: An International Forum*, 10, 57-64.

1.5.3 评价研究

Greene, Jay P. , & Peterson, Paul E. (1996). *Methodological issues in evaluation research: The Milwaukee school choice plan* (Occasional Paper 4). Cambridge, MA: Harvard University, Department of Government.

LaFond, Colette, Toomey, Traci L. , Rothstein, Catherine, Wagenaar, Alexander C. , & Manning, Willard. (2000). Policy evaluation research: Measuring the independent variables. *Evaluation Review*, 24, 92-101.

Toffolon Weiss, Melissa M. , Bertrand, Jane T. , & Terrell, Stanley S. (1999). The results framework—An innovative tool for program planning and evaluation. *Evaluation Review*, 23, 336-359.

Watt, James H. (1999). Internet systems for evaluation research. *New Directions for Evaluations*, 84,23-43.

2

BASIC RESEARCH DESIGN

基础研究设计

第 2 部分强调将基础研究的规则作为所有研究的共同特征，无论研究在性质上是基础的、应用的，还是评价的。它展示了在计划一个研究项目的过程中最初的几个步骤的指南：选择和定义一个社会学的问题，描述问题与某种理论框架之间的关系，形成工作假设(包括零假设和研究假设)，实验或者研究的设计，抽样程序等等。

2.1 社会研究课题申请基本程序

对于社会科学研究的努力来说，在进行实地和实验室的工作之前，应该将某些元素放到行动大纲和计划中去。《研究设计与社会测量导引》的这一部分中所选择的每个单独的说明，都是为了帮助我们计划五个主要的步骤。第3部分（应用和评价研究）及第5部分（收集资料的方法和技术），都提供了许多组对这里所列出的要点的更广泛的和更进一步的说明。

以下将一点一点地说明如何产生一个可接受的研究申请。你甚至可以用下列方法作为帮助你在发展一个研究申请的过程中保持正确方向的清单：

1. 确定和描述社会学的问题
   A. 明确表述研究的问题，必要时对重要的和相关的概念进行定义。
   B. 表明如何通过一定处理或检验来研究这一问题。
   C. 用下述一个或多个标准来描述问题的重要性：
      ①它反映了一个或多个在基本重要性方面特别适时的论点。
      ②可以表明所考虑的问题与一个实践问题相联系。
      ③结果对广泛的受众具有重要意义。
      ④手边的问题与一个有影响的或者重要的总体有关。
      ⑤所完成的工作将会填补研究文献中的空白。
      ⑥所完成的工作将可能概括成有关社会交互作用的更为广泛的原理或者一般性的理论。
      ⑦可以精炼一个重要概念或关系的定义。
      ⑧结果可以对实践的问题有一个更广泛的应用。
      ⑨过去的结果一直是不确定的，需要实验的复制。
      ⑩可以发展出新的或更好的收集和分析资料的工具。
      ⑪增加了收集资料，特别是收集受到时间限制的特定资料的机会。
2. 理论框架
   A. 描述问题或研究问题与理论框架之间的联系。
   B. 展示问题陈述如何与先前的研究相联系，以及先前的研究如何使得当前所提出的众多问题范围缩小。
3. 假　设
   A. 清楚地陈述零假设和研究假设（或备择假设），在理论框架的范围内说明备择研究假设的可行性。
   B. 通过指定显著性程度或Ⅰ类错误的程度来确定你所愿意冒的风险水平。
   C. 用操作性语言来定义概念或变量：

①需要对自变量和因变量进行操作定义。

②说明用于评估这些变量的特定工具,如果可能,在附录中附上例子。

③提供和讨论包括工具信度和效度在内的心理测量学数据。

D. 描述可能的不足及其后果。

4. 实验设计

A. 描述特定的实验设计,要特别关注所有的变量,以及变量是否被检验、被控制或者被忽视。

B. 确认对内在效度与外在效度所存在的威胁,并对如何处理进行说明。

C. 指定统计检验:

①指定检验每一个零假设所采用的置信水平。

②对于给定的研究假设,确定检验的解释力。

5. 抽样程序

A. 描述实验组和控制组。

B. 确定和说明与研究假设相关的总体:解释样本规模是如何确定的以及所选择的样本类型如何。

C. 确定抽取或选择样本的方法。

D. 根据各种样本规模和抽样类型估计相关的收集资料费用。

6. 资料收集

A. 估计资料收集的时间。如果合适,明确指定下列事项:

①向每一个回答者提出的问题数目。

②访问大概需要的时间。

③已建立的时间日程表。

④对访问和结果的预试。

⑤当上述1~4项出现偏差时用于指定适当的时间和地点的备用计划。

B. 对下列适当的访问程序的描述:

①获得信息的方式(例如,通过直接访问、全部的或部分的邮寄、电话,或其他方式)。

②描述对访问员的特别训练。

7. 结果分析

A. 说明分析每一个研究假设的方法:

①所采用的统计软件包。

②通过例子说明表格等等的运用。

③所采用的图表制作技术。

8. 结果的表达与讨论

A. 对资料如何与假设相关进行讨论。

B. 对可替换的其他分析计划进行讨论。

C. 目前研究的不足及其可能的解决办法。

D. 未来研究的思路。

9. 成文、出版或报告计划

A. 申请报告应该符合它所申请的委员会或研究机构的形式要求,通常,这种要求是《美国心理学协会出版体例》(*Publication Manual of the American Psychological Association*. 2001,第5版),或者是《美国社会学协会出版体例指南》(*American Sociological Association Style Guide*. 1997,第2版)。

B. 根据系里和研究生院的要求完成最终的文件。

C. 根据特定的委员会的指南选择一家合适的委员会去申请。

2.2 形成研究问题

研究设计的第一步是从一种令人感兴趣的理论争论中选择一个问题。社会研究的潜在主题范围如同社会行为的范围一样广泛。虽然这一事实并不会使得研究者作出选择变得更为容易,但它却说明了值得探索的事物是一个非常广阔的领域。此外,没有研究者会忽视将他个人的生活以及职业的经历作为导致一个研究问题的思想的来源。作为形成思想的最好的发现经常起始于在散步或者在清晨独处时所找到的线索中,而不是在教室或者实验室中。

问题的选择需要同时考虑到时间、钱以及精力。一个研究者花六个月到一年的时间来寻找一个特定的问题和形成一个可以实际研究的问题,这是十分常见的。进行这项研究可能会花费好几年时间,而要通过出版物来了解这项研究则需要更长的时间。

问题的重要性及其所形成的研究问题的精确性,在为特定领域的知识整体作出贡献中所扮演的角色都不能被低估。明显的问题是:"对于还没有进行的研究来说,这种重要性如何能预见到呢?"在很大的程度上,这是所有的研究者都面临的挑战,也是为什么最初提出的研究问题的质量极为重要的原因。有能力的和有经验的研究者非常熟悉研究文献,他们能够认识到对于更为一般的和理论性的问题来说,下一步需要进行什么样的研究。将所提出的问题置于一种丰富的理论和方法论背景中是十分有效的。

在训练过程中,应该通过识别感兴趣的和潜在多产的问题来对学生(或者任何研究者)进行引导。下列一组一般性要点对于如何发现一个研究的主题(包括硕士论文或博士论文研究活动)以及与此相伴的研究设计提供了建议。

- 从你所修的第一批课程中识别有兴趣的主题(这些课程通常是核心课程或必修课程)。
- 提出问题并将答案并入你在课程论文中已获得的知识中。
- 考虑自己的经历、个性特征以及素质,比如你潜在的专业发展和未来的职业兴趣,与你一同工作的人员的能力和兴趣,你的能力和对与某个特定教授一起工作的期望,以及完成你的工作可利用的资源(钱、设备等等)。

当你继续你的课程论文,同时开始觉得比较适合作为一个有抱负的研究者的时候,记住做下列事情:

- 参加讨论班，通过讨论班你将会经历智力的发展。
- 仔细地考察其他人（特别是高年级研究生）是如何接近研究问题的，为了达到他们的目标，他们在这一过程中采取了哪些步骤。
- 在你自己感兴趣的方面进行小的研究项目（通常是在教师的指导下，作为一项新的研究，或者作为一项与教师正在做的研究项目有关的研究）。
- 与你的导师讨论可能的主题，考察这一领域中有哪些尚未探讨的方面值得研究，前人已有的研究已做了哪些工作，有哪些相关的文献涉及所提出的特定问题。
- 只要可能，尽量参加专业会议。虽然这些会议可能在很远的城市召开，而且你的资源也有限，但是要记住，研究生多年来一直是合住房间，也一起学习（也很有趣）。

这一组建议假定学生认真地对待他们的工作，不是仅仅以一种缺乏兴趣的机械的方式"匆匆写就一篇论文"，而是计划进行一种学术探讨的过程。从一篇论文中或许会产生出几篇公开发表的文章，这些文章将为研究者在这一领域中的声望打下基础，并成为研究者未来发展和贡献的起点。

要成为一名有能力的研究者并无捷径可走，它在各方面都需要大量的时间和实践。在不同背景中经验的增加会使能力得到增强，特别是对于没有经验的研究者来说，最好从对现存研究工作的复制开始。社会科学非常需要这种研究，已获得累积的证据，特别是当先前研究的结果可疑时，学生们可以利用二手资料来丰富研究，同时减少时间和其他资源的花费。

有几本书可以帮助研究新手选择值得研究的重要问题，其中的两本列在下面：*

- Creswell, John W. (1994). Research design: Qualitative and quantitative approaches. Thousand Oaks, CA: Sage.
- Denzin, Norman K., & Lincoln, Yvonna S. (2000). Handbook of qualitative research. Thousand Oaks, CA: Sage.

评价现有研究 2.3

准备任何研究申请的一个完整的部分是对现有文献的评论。在本书的第 4 部分中，你将了解许多有关如何做这一工作的内容。这里，我们的焦点集中在评价现有研究时所应该采用的标准上。这些现有研究构成了我们希望完成的研究的基本原理的基础。作为一个有效的研究者，对任何一项研究的突出特征和缺陷，你都需要一种批判的眼光。

---

* 这两本书最新版的中译本都将于近期由重庆大学出版社出版。

2.3.1 对你评价一项研究的过程和写作文献评论提供帮助的问题

(1)现有研究的评论

- 你的研究中所评论的文献与先前的文献之间的联系是否紧密?
- 评论是否是最近的? 是否有重要的参考文献(那些具有至关重要的概念意义的文献)被漏掉了?

(2)问题与目的

- 你能否理解问题的陈述?
- 研究目的的表述清楚吗?
- 研究目的与所评论的文献的联系紧密吗?
- 是否有一个作为假设建立之基础的概念化的基本原理?
- 对研究的重要性是否有理论的说明?

(3)假　设

- 对研究假设的陈述是否清楚和明确?
- 假设是否陈述了一种清楚的变量之间的联系?
- 假设是否基于某种理论或者基于对相关文献的评论和表述?
- 这些假设是可检验的吗?

(4)方　法

- 是否清楚地定义了自变量和因变量?
- 变量的定义和描述是完整的吗?
- 是否清楚研究是如何进行的?
- 关于因变量是否有足够可靠和有效的数据?

(5)样　本

- 样本的选择是按照一种你认为对总体具有代表性的方式进行的吗?
- 是否清楚样本来自哪里以及它们是如何被抽选的?
- 这一研究中的对象与那些一直用于其他类似研究中的对象的相似程度如何?

(6)结果和讨论

- 作者将研究结果与文献评论相联系了吗?
- 研究结果与假设相关吗?
- 对结果的讨论与结果一致吗?
- 讨论中是否提供了对作者表述的原始假设的结论?

(7)参考文献

- 所列出的参考文献是否是近期的?
- 参考文献的格式是一致的吗?
- 参考文献是完整的吗?
- 所列出的参考文献是否反映了这一领域中最重要的参考文献?
- 论文中所出现的每一处被引用文献是否都出现在参考文献的目录中?

(8)有关报告的总的评论

- 报告写作是否清晰易懂?
- 语言是否是无偏的(非性别歧视的和相对文化中立的)?
- 研究的优点和不足是什么?
- 研究的基本应用是什么?
- 如果要改进研究,应该做些什么?

2.3.2 一个“好的”研究问题的有用和有益的标志

我们如何判断一个研究问题的价值? 提出下述7个问题并将它们应用到你所考虑的主题中去。

- 问题是否涉及一些基本的概念,这种问题所产生的知识是不是累积的并建立在一种现存知识整体的基础上?
- 对研究问题的探究将会得到一种对某些理论规律的检验吗?
- 研究问题允许对所包括的变量进行仔细的说明,同时可以采用最精确的和最合适的方法吗?
- 对研究问题的探究将会通过发现、发展或改进实践的工具、技术或方法得到一种对方法论的贡献吗?
- 研究问题将从所感兴趣的学科或子学科中利用相关的概念、理论、证据和技术吗?
- 这个单一的研究整合到一个计划好的相关研究项目中,将产生比整合前所获得的那些结果更有意义的结果吗?
- 研究问题将对其他研究者的训练提供机会吗?

研究设计的元素 2.4

在为了某种可能的研究考虑而选择一个问题的时候,必须考虑并正式评价整个研究设计以及它的全部元素。表2.1所列出的是在选择问题、具体提问、形成假设以及完成实验的整个过程中必须考虑的研究设计的各个方面。后面几页中对设计中的许多内容提供了指导。研究设计的选择最为重要,因为它影响到研究的全部结果。

表 2.1　决定研究设计特征的重要因素

研究的元素	选　择
基本理论的类型	· 一般理论 · 中层理论(middle-range theory) · 猜想
研究设计	· 预实验的(调查) · 实验的 · 准实验的
接触组织和回答者	· 需要个人的允许 · 需要组织官员的允许
对所研究的社会系统的控制	· 无控制 · 部分控制 · 完全控制
可用的资料类型	· 仅有个案和观察研究 · 仅有定量分析 · 以个案和观察研究作为补充的定量研究 · 其他(历史的、跨文化的等等)
时间维度	· 来自单个社会单个时期的案例(横剖的) · 来自单个社会多个时期的案例(时间序列分析或纵贯研究) · 来自多个社会单个时期的案例(跨文化比较) · 来自多个社会不同时期的案例(比较的纵贯研究)
研究的样本或总体	· 一个群体中具有某种角色的个体 · 成对的相关群体成员(成对物) · 初级群体(30 人或更少) · 次级群体(31 人以上) · 第三级群体(人群、公众等等) · 州、国家或者社会
样本规模	· 一个或很少几个个案 · 小样本(30 以下) · 大样本(30 以上)
资料来源	· 需要研究者收集的原始资料 · 现成的文献或二手资料 · 需要收集的文献或二手资料
收集资料的方法	· 直接观察 · 访问 · 问卷调查 · 测验或其他形式的测量

续表

研究的元素	选 择
自变量的数目	·一个 ·一个以上
因变量的数目	·一个 ·一个以上
测量的水平	·定类的 ·定序的 ·定距的 ·定比的
评价因变量的尺度选择	·具有信度和效度的资料 ·缺乏信度和效度的资料
因变量的特征	·正态分布 ·非正态分布
研究的周期	·短暂的(少于6个月) ·长期的(6个月以上)
完成研究所需要的资源	·需要基金支持 ·不需要基金支持

资料来源:Matilda White Riley. Sociological Research: A Case Approach (New York: Harcourt, Brace & World, 1963)

选择一种研究设计 2.5

社会科学中的经验研究是在各种环境和背景中进行的。任何一个研究项目中,设计背景的选择,对于希望决定某一假设的有效性,以及如何最好地发现接受或拒绝这一假设的证据的研究者来说,通常都是至关重要的。社会现象总是复杂的,就是在最佳状况下对那些导致这种复杂性的相关变量进行控制,也是十分困难的。

因此,现在主要的问题是,“什么样的设计最利于探知所有相关变量之间的联系和因果路径?”研究设计如何着手对问题的回答,将决定研究的未来结果。它将首先决定研究所需要的时间、经费以及其他一些必需的资源。

“起点明确”的总方针应该取代任何其他的设计考虑。每一点努力都应该用于选择设计的背景,使之采用在自变量和因变量两方面都具有大量可变性的总体。一般来说,可变性越强,群体之间的差别就越是可能被察觉。此外,对于任何研究项目来说,保险也是重要的。将个案分析同任何其他的研究设计结合起来具有更好的可靠性。对处于一种分布尾端的极端个案作特别努力的考察,可能会揭示出两极分化的关系,这种关系将提示新的假设、设计和对资料的分析。

表2.2显示的是用于基础研究和应用研究的许多不同的设计的总结。它描述了设计的类型、核心特征以及可能的结果。你会发现其中有许多是相似的。你所选择的

那一种设计(或设计组合)主要是由你所提出的研究问题的性质以及反映这一问题的假设所决定的。

表 2.2　用于基础研究和应用研究的不同的设计

研究设计背景的类型	核心特征	预期的结果
描述性调查 ·横剖研究 例子:美国十年一次的人口普查;Jams A. Davis:《大学生的职业决定》;Peter M. Blau & O. D. Duncan:《美国职业结构》。 ·纵贯研究 例子:Greg J. Duncan & Jams N. Morgan:《5 000 个美国家庭》。	·关注通常由访问或邮寄问卷所得到的资料。其他的资料来源包括官方报告和统计。有时,其他研究者的资料库也可提供合适的数据。 ·要求努力获得所研究样本中的每一个个体的资料。 ·可以产生随时间变化的社会或行为变化情况。	·可以依据类型、频率以及集中趋势的测量来分类的大规模资料。 ·如果总体较大则调查的费用将非常大。最后将产生可以用于分析数量多种关系的数据。 ·标准化的数据可用于对连续的时间间隔进行比较分析,在这种分析中,纵贯研究考察不同年龄组发生的变化,而横剖研究则考察不同年龄组间的差别。
抽样调查 例子:Gallup, Harris, & Roper 的民意测验和公共舆论调查;人口普查局:《当前人口调查》;William H. Sewell & Robert M. Hauser:《教育、职业与收入》。	·仅仅讨论总体中的一个部分。 ·运用抽样方法来产生一个可以精确代表总体的样本。 ·进行假设检验。为了保证有效度,研究者将利用尺度化技术,仔细考虑问卷的用语和表达方式,还要在问卷中包括有关个人背景方面的问题和其他潜在的有用变量。	·数据将用于分析两个变量之间的简单相关关系,以及包括因子分析、矩阵分析、多元判别分析在内的多变量分析。 ·采用合适的参数分析和非参数分析方法来分析定量的和定性的资料。
实地研究 例子:Robert Lynd & Helen Lynd:《中镇》和《转变中的中镇》;August B. Hollingshead:《榆树镇的青年:社会阶级对榆树镇青少年及其重返榆树镇的影响》;William F. Whyte:《街角社会》;Phillip E. Hammond:《工作中的社会学家》。	·主要关注所考察的单个群体、家庭、机构、组织或社区的过程和型式。 ·重点放在社会结构(例如,结构各个部分之间的相互关系以及所发生的社会互动)。 ·希望观察社会互动或者扮演独立角色的人们相互之间的感知和态度。 ·经常使用直接的和参与的观察、访问、尺度化技术。	·所收集的资料能够检验许多采用调查资料检验不了的假设。 ·通过将焦点放在大的总体中的一个小群体上,可以达到更大的控制。适合研究诸如过程、模式、角色、态度、价值等等社会学的产品。

续表

研究设计背景的类型	核心特征	预期的结果
有关个人的个案研究 例子：Elizabeth Eddy：《成为教师：职业地位的转变》；Erwin O. Smigel：《街巷律师》；W. F. Cottrell：《铁路员工》。	· 通常涉及对所考察的现象的一个单个的例子进行相对透彻的分析。 · 研究者对个体进行访问，或者研究其生活史文献，以获得理解其行为的眼光。 · 试图发现独特的特征和某个特定阶层的人所共同具有的特点。 · 个案被分成不同的类型以便发现一致性。	· 资料的聚集可以提供观察关系状况和影响因素的眼光。 · 可以识别个性和社会化的过程。 · 可以发现和检验概念。 · 个案可以进行编码，并应用统计检验来了解变量之间的相关。
调查与个案研究相结合 例子：E. W. Burgess & Leonard S. Cottrell：《预测婚姻的成功与失败》；Alfred C. Kinsey 及其同事：《人类男性与女性中的性行为》。	· 调查方法论与对特定个案的研究相结合，以便对某一相关模式中的联系进行肖像化的初步描绘。并通过个案研究来解释并展现不同的过程和模式。 · 在调查之后选择个案，以揭示出那些在某个变量上或高或低的情况，以及那些显示重要特征的情况。	· 与更加深入地揭示个人社会化过程和模式的资料相伴随的关系。 · 可获得两种资料库：统计资料和个案分析资料。
预测研究 例子：Sheldon Glueck & Eleanor Glueck：《预测违法行为与犯罪》；Paul Horst：《个人成就预测》（见 Paul Wallin：《从个案研究中预测个体的行为》）。	· 目的是在参与一项具体活动之前，估计个体表现的水平。 · 从一个总体中寻找作为预测诸如婚姻的成功或失败、假释成功的程度、早期发现潜在的违法行为、学业成功、犯罪行为这样一些结果的基础因素。 · 经常考察一种不连续的因变量，比如，稳固的婚姻与破裂的婚姻（离婚）、守法行为与犯罪行为、违法的与不违法的男孩与女孩、专业有成就者与专业无成就者。	· 决定许多因素之间的关系和一种预测的标准。 · 在预测表的建构中对所选择的因子赋予权重。 · 利用预测表进行预测。

续表

研究设计背景的类型	核心特征	预期的结果
控制试验(主要类型有实验室实验、“自然”实验,以及实地实验)		
· 实验室实验 例子:Robert Bales:《个性与人际行为》	· 研究者创造出一种他所希望的带有确切条件的情境,在这种环境中,他控制某些变量,同时操纵另一些变量。研究者在一种将其他因素控制起来,或者在合适的情况下将其忽略的情形中,观察和测量自变量对因变量的影响。	· 所发现的关系可以更精确地看作是控制“干扰”变量的结果。
· “自然的”实验(例如横剖的或不包括过去事实的) 例子:F. S. Chapin:《社会学研究中的实验设计》	· 研究者在正常的社区背景中,利用某种正在发生的变化,以一种实验的设计研究它们的影响。 · 对一个群体的人实施某种处置或社会项目,把他们个人的调整与没有给与这种处置和社会项目的另一个群体的人进行比较。 · 当通过配对使所选择的因素保持一致时,群体的配对可以使两个群体相似。	· 在控制的条件下发现和揭示因果关系。有可能作出更严格的陈述。表明社会处理或社会项目效度的增加。
· 实地试验 例子:J. G. Miller:《社会过程中的实验》	· 包括研究者对条件的操纵,以便决定可能的因果关系。 · 将自变量和因变量的最大的变化放到设计的结构中去。 · 建立起实验组和控制组,使那些被认为对所研究的关系有干扰的因素保持不变。	· 自变量(处置)可以广泛变化,可以发现一个敏感的或者确定的标准变量。 · 配对资料提供了最有力的可能的控制。 · 可以推断一个高度可信的因果模型。

参考文献

Bales, Robert. (1969). *Personality and interpersonal behavior*. New York: Holt, Rinehart, and Winston.

Blau, Peter M., & Duncan, O. D. (1967). *The American occupational structure*. New York: Wiley.

Burgess, E. W., & Cottrell, Leonard S., Jr. (1998). *Predicting success or failure in marriage*. London: Routledge/Thoemmes Press.

Chapin, F. S. (1974). *Experimental designs in sociological research* (Rev, ed.). Westport, CT: Greenwood.

Cottrell, Fred W (1971). *The Railroader.* New York: Irvington.

Davis, James A. (1965). *Undergraduate career decisions: Correlates of occupational choice.* Chicago: Aldine.

Duncan, Greg J., & Morgan, James N. (Eds.). (1980). *Five thousand American families: Vol. 8. Analyses of the first eleven years of the Panel Study of Income Dynamics.* Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan.

Eddy, Elizabeth. (1969). *Becoming a teacher: The passage to professional status.* New York: Teachers College Press.

Glueck, Sheldon, & Glueck, Eleanor. (1959). *Predicting delinquency and crime.* Cambridge, MA: Harvard University Press.

Hammond, Phillip E. (Ed.). (1964). *Sociologists at work: Essays on the craft of social research.* New York: Basic Books.

Hollingshead, August B. (1948). *Elmtown's youth: The impact of social classes on adolescents.* New York: Wiley.

Hollingshead, August B. (1975). *Elmtown's youth and Elmtown revisited.* New York: Wiley.

Horst, Paul. (1941). *The prediction of personal adjustment: A survey of logical problems and research techniques, with illustrative application to problems of vocational selection, school success, marriage, and crime.* New York: Social Science Research Council.

Kinsey, Alfred C., & Associates. (1998). *Sexual behavior in the human female.* Bloomington: Indiana University Press.

Kinsey, Alfred C., Pomeroy, Wardell B., & Martin, Clyde E. (1998). *Sexual behavior in the human male.* Bloomington: Indiana University Press.

Lynd, Robert, & Lynd, Helen. (1975). *Middletown: A study in American culture.* New York: Harcourt, Brace.

Lynd, Robert, & Lynd, Helen. (1982). *Middletown in transition: A study of cultural conflicts.* New York: Harcourt Brace Jovanovich.

Miller, James Grier. (Ed.). (1950). *Experiments in social process: A symposium on social psychology.* New York: McGraw-Hill.

Sewell, William H., & Hauser, Robert M (1975). *Education, occupation, and earnings: Achievement in the early career.* New York: Academic Press.

Smigel, Erwin O. (1969). *The Wall Street lawyer, a professional organization man?* Bloomington: Indiana University Press.

U.S. Department of Commerce, Bureau of the Census. (various years). *Current population survey.* Washington, DC: Author.

Whyte, William F. (1993). *Street corner society: The social structure of an Italian slum.* Chicago: University of Chicago Press.

关于研究设计与研究问题选择方面的推荐读物

Allwood, Carl Martin, & Barmark, Jan. (1999). The role of research problems in the process of research. *Social Epistemology*, 13, 59-83.

Bailey, Kenneth D. (1999). *Methods of social research.* Collingdale, PA: Diane.

Bechofer, Frank, & Paterson, Lindsay. (2000). *Principles of research design in social sciences.* New York: Routledge.

Garratt, Dean, & Hodkinson, Phil. (1998). Can there be criteria for selecting research criteria? A hermeneutical analysis of an inescapable dilemma. *Qualitative Inquiry*, 4, 515-539.

Grunow, Dieter. (1995). The research design in organization studies: Problems and prospects. *Organization Science*, 6, 93-103.

Hazen, Dan, Horrell, Jeffrey, & Merrill-Oldham, Jan. (1998). *Selecting research collections for digitization.* Washington, DC: Council on Library and Information Resources.

Mitchell, Mark, & Jolley, Janina. (1995). *Research design explained.* Fort Worth: Harcourt College Publishers.

Reiser, Stanley Joel, & Bulger, Ruth Ellen. (1997). The social responsibilities of biological scientists. *Science and Engineering Ethics*, 3,137-143.

Sasaki, Masamichi. (1995). Research design of cross-national attitudinal surveys. *Behaviormetrika*, 22, 99-114.

Tabachnick, Barbara G., & Fidell, Linda S. (2000). *Computer-assisted research design and analysis.* Needham Heights, MA: Allyn and Bacon.

2.6 科学是怎样建立的

本书阐述了行为研究者眼中的科学经典作品。同时,它们应当被用作在任何科学探索研究中以指明方向的一系列路标和确定其固有的潜在困难的依据。这就像是把研究者始终持有的,并运用在工作和生活中,引导其研究的方方面面的哲学融会贯通起来一样。

以下三篇短文都是在你开始考虑你所从事的研究主题和职业活动时非常值得一读的材料。

概念定义和理论建构的重要性

Gordon J.DiRenzo

科学调查研究力求通过建立普遍原理解释在我们经验世界中所研究的现象,并且人们希望科学能够预言这类现象。科学的诸多原则最终被表述成我们所知晓的理论。为了解释具有真实性的事实,科学家需要一个组织好的概念系统。一门“无概念科学”是不可能存在的——就像任何形式的没有概念的理性行为一样是不可思议的。然而,说概念对于科学是不可缺少的只不过是为了假定或让问题变得可能。在这里,指的是所需的科学元素的定义和建构。

最初在社会调查中,对于现象的描述凭借一种非技术性的语汇表述出来。一门学科的增长进步立刻包括了关于概念和相应术语的思索系统的发展。这些系统性的思索或多或少都是有些抽象的。然而,即使经过了几十年的定义和再定义,许多科学中

的基本的术语还远不能够与普遍所接受的定义区分开来。这在某一特定学科之内与在这一学科之外一样普遍。例如,如果给行为科学中的三个关键概念命名,就会有许多对于"社会"、"文化"和"个性"的指称符号。科学的技术性的概念是怎样被引入的,并且它们在科学的进程中怎样发挥作用,在这里成为了中心问题。

概念定义和理论建构携手前进成为科学研究统一进程中必不可少的环节。概念分析是一个使科学探究的复杂进程在理论上达到顶点的阶段——一个基本的必需的阶段。伴随着行为科学中对"理论"这一术语越发独特的理解,概念因此成为理论或理论系统中不可缺少的元素。概念要素越精确和精炼,那么理论也就越精确和精炼。

我们针对自己所阐述的问题对于科学调查的所有领域来说都是一个基本的问题。

资料来源:Gordon J. DiRenzo(Ed.)(1967),*Concepts, Theory, and Explanation in the Behavioral Sciences* (New York: Random House),使用获得授权。

科学方法运用中潜藏的假设

Gideon Sjoberg & Roger Nett

成为科学方法应用基础的一组最简化的假设(通常不被表述)包括:①存在一个确定的事件循环的顺序。②这种知识胜过无知。③存在一条基于知觉印象、存在于科学家和"客观外在现实"之间(被称作"经验假设")的沟通纽带。④在物质和社会次序范围内存在因果关系。⑤有特定的"观察者"假设。这一假设又包括:a. 观察者是被其试图改进人类状况的愿望驱使而去获取知识的;b. 观察者有能力从概念上将观察资料联系起来并对事件进行归因;c. 社会将会支持研究者追寻所要的知识。这些被科学家或多或少认为理所当然的假设,在最后的分析中被广泛地理解为"功能性虚构"。它们在获取知识上的有效性是其首要的存在理由。

关于"自然"世界中次序的假设

科学,就其对所在范围内的概括和预测来说,依赖于其所研究的物质世界和社会世界中所存在的某种次序的程度。科学不能将其描述成规则性的表现,它必须定义为一些可描述的与规律性的一种背离。这种推断假设事件是按照某一特定的维度被排序的。固然,所有的知识体系都建立在宇宙中对于规律的假设基础之上,但是这种假设对于科学来说比对于其他知识体系可能具有更重要的意义。毕竟,科学家花费其大部分的时间对不同种类的相对一致性进行区分,并且将这些种类与其他种类联系起来。甚至在一次迅猛的变化中,革命系统存在一定程度的规律,并且变化本身显示了能够被描述和分析的不同模式。

对于规律的假设引导社会科学家(如果他们仍是科学家的话)避免历史循环论。那些以极端形式拥护历史循环论者位置的人,设想每一种文化系统都必须作为一个单独的整体被研究。而且,他们还认为跨越文化无规律可循。当然,甚至历史循环论者也承认有各种类型的统一体存在。因为他们认识到了每一个系统都有它自身发展的

规则。

规律的概念与"自然宇宙"的概念密切相联。在对科学史的描绘中我们发现,当学者们能够用自然主义的术语来构想物质和社会环境时,也即是如同精神领域中的因素独立地发生作用时,一次重大的突破就产生了。在现代人发展出操纵和正确控制物质与社会范畴内方方面面的手段的过程中,这是基本的一步。

证据表明,如果像19世纪的乌托邦思想家那样将按科学方法建立的知识同智慧相混淆是错误的。智慧包含了合理的道德方向、良好感觉的训练和对是否值得做的问题的区别能力。

科学的方法(在较狭窄的意义上)除了促进科学的目标外,并没有告诉我们如何经验地运用经过检验的知识。然而,通过更多地利用被经验证实有效的知识和较少地利用未经证实的以及常常是其他认识论的肤浅的知识,人类的认识将得到提高。

资料来源:Gideon Sjoberg & Roger Nett(1968),*A Methodology for the Social Researcher*(New York,Harper & Row),使用获得授权。

研究者的两难境地与行为科学的特殊性

Abraham Kaplan

两难境地

在进行研究的过程中,我们不断地屈从于来自相反方向的拉力,例如:搜寻资料或者形成假设,建构理论或者进行实验,关注于一般规则或者是个别案例,进行质量研究或者分子研究,进行综合或者进行分析。实际上被告知两者都必须做是没有多大帮助的。在特定的有疑问的情形约束下,这些是真正的两难困境。但它们是一种被认为是存在主义的两难境况:并不是对于一些特别历史情况有特殊性,而是对于真理的追求有本质意义。我们并不是两害之中取其轻和遵从于不良的结果。存在主义两难困境造成的问题根本不能得到解决而只能被处理,也就是说我们要学着让它与我们共同存在。"我们需要的是勤奋的工作者和经验主义,而不是灵感。"有很好的理由要求这样做。但同等重要的理由可以被用来支持相反的观点。事实是,我们需要所有我们所能获得的东西。事物的这种情形对于行为科学来说绝不是罕见的。就我所观察到的情况而言,它的方法论与其他无论任何一门科学没有什么不同。当谈及"科学方法"时,这种一致性是预期到了的,那么我强烈赞成使用它。

行为科学的特殊性

因此,行为科学的与众不同之处是,什么成为了它的基本主题。主题允许的或者需要的技巧都仅仅是派生出来的。如果要求有一些关于这个主题的单一判别式的话,我相信最广泛适用的一种是由C.W. Morris提出的"符号"的运用。行为科学所处理的是那些由符号,或者至少由意义作为基本部分的过程。"意义"被解释的范围有多

宽,或者在最宽泛的解释中它包含了多少动物行为,在这里都是不必烦扰我们的问题。毫无疑问的是,无论我们如何界定行为科学的界限,它还是会超出自身范畴涉及生物学。但是,这种困难比起方法上的困难(对于基金会、图书管理员和院长们来说)更容易管理控制。

这里非常明显的是,行为科学的资料不是纯粹的动作而是行动——即这种行为是在一种赋予其意义或目的的观点下被执行的行为。通俗地说,我们区分行为对于行为者(或者对于他人,包括对他做出反应的我们自身)的意义和对于我们作为科学家的意义是至关重要的。我分别将它们称为行动意义(act meaning)和行为意义(action meaning)。我一会儿将回到这个特性上。当前,我们应该注意到,行为科学包含在解释的双重过程中,并且,这正是它的技巧如此与众不同的原因所在。行为科学家必须首先达到行动意义,即解释导致一项特定行为的片断代表了什么意义;然后他就必须搜索被解释的行为意义,以及它与其他行为和情境的相互联系。他必须首先看到作为进行选举的行为的记选票或者操作机器的行动,然后再继续他关于选举行为的研究。

现在,尽管对于行动意义的解释通常包含了特殊的技巧,这些都还是服从于相同的用以统治行为意义解释的方法论标准的(并且其他科学也是这样)。我们不依靠任何特殊的努力(用我们自己的语言)解释言语行为——确实,通常对于行为作为行为本身没有任何意识(我们只听到说了些什么,却并没有听到具体的言语)。然而,每一种解释都是一个假设,每一个答复都是一次实验——我们最终可能已经误解了其意义。当碰到解释外语并且大体上解释另一种文化的各种形式时(尽管它们本质上并没有什么不同),情况就变得更加清楚了。事物这种状态的一些含义将与行为科学中的"verstehen"角色相联系而被探究。我在此要强调的观点是,行为科学家试图理解行为的感觉就如同物理学家试图去了解核进程一样。差异并不在于有两种理解,而是对于行为科学家而言有两种不同的事物要去理解。例如,一个精神病学家需要了解为什么一个病人发出某种噪音(为了告诉他的临床医生他有多恨他),并且他为什么要说出他做的事(因为他还没有完成传送的过程)。显然,因为我们是我们人类本身,所以我们对于理解噪音有自己特殊的方式。但出于同样的原因,我们对于解释光波也会有特别的方式,但对于无线电波需要许多其他的技巧。关键是就连我们所看到的也并不一定是永远可以相信的。每一种技术都服从于实证检验,并且相同的标准都运用到它们当中。

资料来源:Abraham Kaplan(1964), *The Conduct of Inquiry* (San Francisco:Chandler), pp. 30-33,使用获得了授权。

关于科学方法与科学理论的推荐读物

Carey, Stephen S. (1997). *A beginner's guide to scientific method.* Belmont, CA: Wadsworth.

Gruender, David. (1998). Values and the philosophy of science. *Protosociology*, 12, 319-333.

Gustavsen, Bjorn. (1996). Is theory useful? *Concepts and Transformation*, 1, 63-77.

Klovdahl, Alden S. (1995). Levels of protection: Confidentiality in network research. *BMS, Bulletin de Methodologie Sociologique*, 48, 120-132.

Lee, Jeffrey A. (1999). *The scientific endeavor.* Menlo Park, CA: Benjamin Cummings.

Longhin, Luigi. (1998). The objectivity of scientific knowledge in the intersystemic approach. *Fenomenologia e Societa*, 21, 74-90.

Nola, Robert, & Sankey, Howard. (1999). *After Popper, Kuhn and Feyerabend: Recent issues in theories of scientific method.* Norwood, NJ: Kluwer Academic.

Poole, Darryl G., & Pickard, Ruth B. (1986). *Disciplinary change and sociological practice: An historical and integrative view.* Paper presented at the meeting of the American Sociological Association.

Sklar, Lawrence. (Ed.). (1999). *The philosophy of science: Vol. 2. The nature of scientific theory.* New York: Garland.

Taylor, Jan. (1987). Retrieval and reconstruction: A sense of purpose for the social sciences. *New Zealand Sociology*, 2, 98-111.

Weinberg, Steven. (1995). The methods of science... and those by which we live. *Academic Questions*, 8, 7-13.

2.7 社会学理论对经验研究的影响

社会科学中极少有研究者会怀疑他们的基础研究和经验研究间有着重要的、复杂的和几乎共生的关系。例如，Robert K. Merton 在这一部分之后的摘录中，描述了理论对于经验研究的影响，并且指出："在一定意义上，定向研究（directed research）的概念意味着，经验调查是极有组织地进行的，一旦发现经验具有一致性时，它们便会对一个理论系统产生直接的影响。"

我们现在关注 Merton 所阐述的理论的作用。研究者必须时常建构能将假设同一项更包容、更大的理论相联系的中层或微观理论。基于同样的关系，Zetterberg（1954）写道：我们的目的是创造一般理论和微观理论以帮助确定和产生简单易行的研究问题。这些微观理论"将部分地由微观理论的途径所构成，支持一项微观理论的实验证据将在该微观理论为宏观理论的一个特例时同样支持相应的宏观理论"（p. 15）。

其他学者也从不同的学科角度认识到理论与经验资料关系的重要性。诺贝尔经济学奖得主 Milton Friedman 已经列出了以下具有重要影响理论的标准：

> 一项理论"越简单"，在现象的某一特定领域中做出预测需要的初始知识越少，那么它就越是硕果累累。并且，该理论对于结果的预测也越为精确，理论做出预测的范围也越广。另外，它还能为进一步深入研究提供更多的附加线索……对于一个假设正确性仅有的相应测试是将经验与预测进行对比。（Friedman, 1953, p. 10）

参考文献

Friedman, Miltion. (1953). The methodology of positive economics. In *Essays in positive economics.* Chicago: University of Chicago Press.

Merton, Robert K. (1957). The bearing of sociological theory on empirical research. In *Social theory and social structure.* Glencoe, IL: Free Press.

Zetterberg, Hans L. (1954). *On theory and verification in sociology.* New York: Tressler.

社会学中的经验概括

Robert K.Merton

通常来讲社会学理论的目标是获得关于社会一致性的论述。对于这个简略的论点,需要加以详细的说明。有两种关于社会一致性的论述,它们在孕育理论方面有着显著的区别。第一种是经验概括:一个孤立的命题概括了观察到的两个或两个以上变量间关系的规律。[1]在社会学文献中有许多这样的概括还没有被吸收进社会学理论中来。因此,恩格尔关于消费的"定律"或许可以作为具体的例子来说明。Halbwachs 关于体力劳动者比属于同一收入水平的白领雇员的人均食品消费要大的发现也可以作为例子说明。[2]这样的概括或许不是很精确,但是这并不影响它们在探索结构中的逻辑位置。对于一个美国城市的样本,Groves-Ogburn 发现,"一般来说,在那些从事制造业的人数百分比较高的城市中,已婚年轻人的比率也相对高一些",这种关系已被用一个方程式来解释它们之间的相关程度。尽管这种类型的命题在经验研究中是最基本的,但是,一堆杂乱的命题只能为社会学学科提供粗糙的材料。理论的任务以及通向理论的经验研究的定位,首先在于弄清楚各种一致性对相互关联的命题发生的作用。定向研究的概念意味着,在一定意义上,[3]经验调查是极有组织地进行的,一旦发现经验具有一致性时,它们便会对一个理论系统产生直接的影响。只要理论是在指导下进行的,在得到发现之前,就可以提出这个发现的基本原理。

社会学理论

社会学概括的第二种类型,被称为科学定律。它和第一种有明显的不同,因为它是一个可以从理论中导出的恒定陈述。在社会学领域中,这种定律的缺乏或许反映了当前理论和经验研究明显的分离。尽管有很多著作论述社会学理论史,也有大量的经验研究,但社会学家们(包括本书作者)仍然可以在不引证一个完全满足这些标准的例子的情况下来讨论社会学定律的逻辑准则。[4]

与这些标准相接近的东西并不完全缺乏。为了展示经验概括和理论之间的关系,以及为了明确理论的功能,考察一个熟悉的案例或许会有所助益。在这个案例中,经验概括已经被结合进了一个实质性的理论之中。那就是,在各种对象的总体中,天主教徒的自杀率要低于新教教徒。这一个案长期以来一直作为统计的一致性而被确立。[5]在这种形式中,一致性提出了一个理论问题。这种一致性只是构成了一个经验的规律性,如果它能够从一系列其他的命题中推导出来——这正是迪尔凯姆为自己确立

的一个任务——那么它就可能成为有意义的理论。如果我们用正式的方式重述他的理论假设,他的理论分析范式就变得清晰了:

- 社会内聚力为承受剧烈压力和焦虑的群体成员提供支持。
- 自杀率是人们承受的压力和不能减缓的焦虑的函数。
- 天主教徒比新教徒具有更大的社会内聚力。
- 因此,可以预期天主教徒的自杀率要低于新教徒。[6]

此案例有助于确定经验概括在与理论的关系中的位置,也可以说明理论的某些功能。

1. 它指出,理论相关性不是天生就存在或者不存在于经验概括之中,而是在概括通过高度抽象地概念化之后(天主教教义—社会内聚力—缓解忧虑—自杀率)才会出现。而这种高度抽象又被包含在更为一般性的关系陈述中。[7]最初作为一种孤立的一致性的东西,被重新作为一种关系,这种关系不再与宗教归属和行为有关,而是与有着某种概念化特征(社会内聚力)和行为的群体有关。这样,最初的经验发现的范围被可观地延伸了,同时,一些看似相异的一致性也似乎变得有关联了(因而,已婚人群和单身人群之间的自杀率差异也可以从这个理论中推导出来)。

2. 一旦从一系列相关的命题中推导出一致性的理论相关性,我们就可以为理论和研究发现提供双重积累。自杀率差别的一致性增加了这组命题的确定性,正是从这组命题中,导出了自杀率差别的一致性以及其他一些一致性。这是系统理论的一个主要功能。

3. 由于这种经验一致性并不允许从自身抽出不同的结果。因而在远离自杀行为的领域中它的再形成会带来多种结论。例如,在对着迷性行为、病态偏见以及其他不适应性行为探究时,发现这些行为也和没有足够的社会内聚力有关。[8]因此,把经验一致性转换为理论陈述,并且通过连续不断的探索应用,就可以丰富研究成果。

4. 通过提供一个基本原理,理论可以导入一种预测的基础,这种基础比只是凭着先前观察所得到的趋势进行经验推断要可靠得多。因此,如果一个独立的测量表明一群天主教徒中的社会内聚力在降低,那么就可以推论出这个群体的自杀率将有上升趋势。而缺乏理论的经验主义者则没有其他的办法,只能在推测的基础上来预测。

5. 先前所列举的几项功能已经预示了理论的一个更深远的特征,它并不完全符合迪尔凯姆的公式,同时,它引出了一个至少到目前为止特别困扰社会学理论的一般性问题。如果一个理论是多产的,那它必定是充分精确的从而是确定的。精确性是可检验性标准的一个主要的元素。当我们考虑精确性在训练有素的研究中的逻辑地位时,社会学中那种只要有可能就使用统计数据来控制和检验理论推论的普遍的压力就有了一种合理的基础。

从一个理论中所能导出的推论(预测)越精确,适合这些预测的替换假设的可能性就越小。换句话说,精确的预测和数据可以帮助减少断言结果的逻辑谬误的研究与经验间的关系。[9]众所周知,从一个理论导出而又被证实了的预测,并不能证明或说明那个理论,它们只是提供了一种对证明的测量,因为,从不同理论体系中抽出的替换假设能够说明已预测到的现象,这种情况总是可能的。[10]但是,那些接受由观察所证明的精确预测的

理论，具有战略重要性，因为它们为我们在相互竞争的假设中作出选择提供了一种初步的基础。换句话说，精确性增强了接近一个“决定性的”观察或实验的可能性。

一个理论的内在凝聚力也具有许多相同的功能，因为，如果被经验证实的多种多样的结果来自于同一个理论体系，这就会减少相互竞争的理论都能适当地说明同样的资料的可能性。比起那种既不同又不相关的假设来，整体性的理论提供了一种对确定性的更大的测量，因此能积累起更加有分量的证据。

寻求精确性和逻辑内聚力的双重压力，可以导致没有收益的行动，特别是在社会科学领域中。任何程序都可以被使用或者被滥用。不惜一切代价过早地坚持精确性，可能会扼杀富有想像力的假设。为了进行有结果的测量，可能会导致对科学问题的重新阐述，使得后来的材料并不受到手中原来的那个问题的影响。[11]在寻求精确性的过程中，要特别注意不能在无意中使重要的问题偏离原先的观点。同样的，逻辑一致性的压力也常常引起文字之争以及无结果的理论化，因为包含在系统分析中的假设时常远离经验的对象或者陷于高度抽象之中而不能够进行经验探索。[12]但是，探索的这些标准的合理性并不会因为这种滥用而受到损害。

注 释

1 正如 Dewey 所指出的，术语“经验的”这种用法是很常见的。在这个上下文中，“‘经验的’意味着：一个具有经验推论的特定命题的主题，仅仅只代表可以被重复地观察到其存在的一组特征的统一联合体，人们反复地观察这些特点的存在，但不能理解为什么这种联合体会出现，也没有一个理论能说明它的基本原理。” John Dewey，Logic：*The Theory of Inquiry*（New York：Henry Holt 1938），p. 305.

2 参见 C C Zimmerman 所总结的这种一体化的大量选集：Consuption and Standards of Living（New York：Van Nostrand，1936），55ff

3 “在一定意义上”，如果仅仅因为它使获得丰富的新发现的可能成为徒劳，研究将完成局限于检验预定的假设。产生于研究过程中的各种预感，对于一个更广阔的理论系统来说，或许没有直接明显的意义，但最终它们会导致经验一致性的发现，这种经验一致性以后可以结合进一个理论。例如，在政治行为社会学中，最近已经证实：投票人受到的社会交叉压力的数量越大，他们在总统选举中所表现出的兴趣越小（P F. Lazarsfeld，Bernard Berelson，and Hazel Gaudet：*The People's Choice*，New York：Duell，Sloan & Pearce，1944，pp 55-64）。这个发现可以引发对政治行为进行系统研究的新路线，尽管它尚未被结合进一种一般性的理论。而当最初对研究进行表述时，这一发现是完全不曾料到的。富有成效的经验研究不仅从理论上检验了导出的假设，它还产生了新的假设。这可以被称作研究中的“偶然发现珍宝的才能”成分，即由于机遇或睿智，发现事先并未寻求的有效结果。

4 例如，参见 George A Lundberg 的讨论：“社会科学中的定律概念”，《科学哲学》，1938，pp. 189-203。这篇文章肯定了这种定律的可能性，而没有举出相关的案例说明。K D. Har 的著作《社会定律》（教堂山：北卡罗来纳大学出版社，1930）并没有履行其题目中所暗含的诺言。一些社会科学家讨论了获得社会定律的可能性，他们发现难以举出例子。Herbert Blumer，《对 Thomas 和 Znaniecki 的在欧洲和美国的波兰农民的评价》，（纽约：社会科学研究委员会，1939），pp 142-150。

5. 当然，这一陈述假定教育、收入、民族、城乡居住地和其他一些可能造成这一发现虚假性的因素都是一致的。

6. 我们不必考察这个说明的其他方面,例如,①对迪尔凯姆的解释中所暗含的前提作出的充分陈述的程度;②将这些前提作为问题而不是作为已知条件的辅助性理论分析;③理论解释潜在的无限回归被终止于某一点而不是另一点的根据;④在导入诸如社会内聚力这样的不能直接测量的中介变量的过程中所包含的问题;⑤前提被经验地肯定的程度;⑥由这种说明所代表的较低的抽象程度;⑦迪尔凯姆从这组同样的假设中导出几个经验概括的事实。

7. Veblen 中肯地指出:“所有这些看起来是在承受琐碎之苦。但是,任何科学研究都不得不使用的资料在某些方面是琐碎的,而在另一些方面则可能是重要的。”Thorstein Veblen,《科学在现代文明中的位置》(纽约,1961),p. 42。

8. 参见,例如,Elton Mayo,《一种工业文明的人类问题》(纽约:麦克米兰,1933),第 113 页和书中各处。如作者所证明的,由 Whitehead,Roethlisberger 和 Disckson 在有关工业士气的研究中所用的理论框架也部分地来自迪尔凯姆的系统阐述。

9. 当然,“通过预测来证明”的范式存在逻辑谬误:

如果 A(假设),则 B(预测)。

观察到 B。

因此,A 是真的。

因为除了包含正式的标准以外,这种范式对科学研究的干扰并不大。

10. 作为一个相关的个案,考虑一下不同的理论家曾经预言过本世纪中叶的战争和大规模的血腥冲突。例如,Sorokin 和一些马克思主义者以大相径庭的理论体系为基础提出了这种预测。如果仅仅根据观察到的事实与二者相一致的话,大规模冲突的实际爆发在本质上并不能使我们在这些不同的分析计划之间作出选择。只有当预测是如此详细地被说明,如此的精确化,使得实际发生的情况与一种预测相一致,而与另一种预测不一致,才能进行一项决定性的检验。

11. Stuart A. Rice 评论了公共舆论研究中的这种趋势。参见:Louis Wirth 等编《十一个二十六:社会科学研究十年》,(芝加哥:芝加哥大学出版社,1940),p. 167。

12. 在经济学领域,正是这种实践被 Walker 称之为“理论枯萎”(theoretic blight)。E. Ronald Walker,《从经济学理论到政策》(芝加哥:芝加哥大学出版社,1943),第 4 章。

资料来源:Robert K. Merton(1957),“The Bearing of Sociological Theory on Empirical Research,” in *Social Theory and Social Structure* (Glencoe, IL: Free Press), pp. 95-99.

2.8 架起理论语言与研究语言之间的桥梁

无数的导师告诉他们的学生,每一个字都有其含义,学生应该说出他们的意思,并且应该知道他们所说的话的含义。当人们谈及我们大家都不得不涉及的研究语言和相关术语时,情况就不再如此。特别是当它面对理论以及探索理论的研究时,语言就扮演了一种特别重要的角色,这一点正如下面提供的 Blalock 的有关讨论。

非实验研究中的因果推断

Hubert M.Blalock,Jr.

1. 科学方法的内在特征,使得理论的语言与研究的语言之间存在着差距。因果推断属于理论的层次,而实际的研究仅仅只能建立起共变与时间顺序。

2. 作为一种结果,我们实际上从来就不能经验地说明因果定律。甚至在实验的条件下也是如此。包括“如果……就”形式的假设陈述在内的因果定律只是科学家的工作假设。

3. 人们承认,因果思考完全属于理论层次,同时,因果定律从来就不能被经验所说明。但这并不意味着,从因果关系的角度进行思考以及发展具有应用性并可以间接检验的因果模型是无用的。在与这些模型打交道时,利用整个系列的不可检验的简单化假设是必须的,这样,即使当一个已知的模型产生出正确的经验预测时,这也不意味着它的正确性可以被说明。

现实,或者至少是我们对现实的预测,无可否认的是由正在发生的过程所构成的。没有两个事件是完全重复的,在不同的时间点上,任何客体或有机体都不会完全保持其形式一点不改变。[1]可是,如果我们曾经理解了真实世界的性质,我们就必须将事件当成好像是重复的,好像客体真的具有在一个时期中保持恒定不变的性质那样去思考和行为,无论这个时期多么短。除非我们允许我们自己作出这种简单的假设类型,否则将永远不能超出对简单的和惟一的事件进行概括。

4. 我们所强调的是,无论设计多么精心,总必须作出某种简单化的假定。特别是,在某些点上我们必须假定,混淆某些因素的影响可以忽略不计。随机化帮助我们排除了一些这样的变量,但是,这种特定的简单化假设类型总是存在一些问题。我们希望强调这种现实,以便强调在实验的基础上作出因果推断的逻辑与在非实验设计的基础上作出因果推断的逻辑之间的基本相似性。

注 释

1. 在 Karl Pearson 的经典著作《科学的语法》(*The Grammar of Science*,NY:1957,5)中强调了这一特殊点。

资料来源:Hubert M. Blalock, Jr. (1964),Causal Inferences in Non-Experimental Research (Chapel Hill:University of North Carolina Press),pp. 172-173,6-7,26,使用获得了授权。

关于社会理论的推荐读物

Jacobsen,Chanoch,& Bronson,Richard. (1995). Computer simulations and empirical testing of sociological theory. *Sociological Methods and Research*,23,479-506.

McLennan,Gregor. (1995). After postmodernism—back to sociological theory? *Sociology*,29,117-132.

Ritzer,George. (1999). *Modern sociological theory.* New York:McGraw-Hill.

Ritzer, George. (2000). *Classical sociological theory* (3rd ed.). Boston: McGraw-Hill.
Wallace, Ruth A., & Wolf, Alison. (1999). *Contemporary sociological theory: Expanding the classical tradition*. Upper Sddle River, NJ: Prentice Hall.
Zeitlin, Irving M. (1996). *Ideology and the development of sociological theory*. Upper Saddle River, NJ: Prentice Hall.

2.9 判断假设用途的标准

一个假设如果不能被检验,它就是无用的。在下列部分中,William J. Goode 和 Paul K. Hatt 展示了用一组标准一步一步地评价假设的方法。特别应注意,他们对假设要与一个理论相联系的强调。

为了完善我们的思考,我们也应注意到假设检验结果的证据。Zetterberg 陈述了接受一个工作假设的三条标准,即:

- 所发现的经验资料与工作假设的预测一致。
- 我们在某种概率的范围内质疑零假设。
- 我们已经证明针对所检验的假设的替换假设是错误的。

社会研究方法

William J. Goode & Paul K. Hatt

1. 假设必须概念清晰。概念应该清楚地定义,如果可能,还应该是可操作的。同时,它们的定义还应该是普遍可接受的,以便交流。

该做什么:澄清概念含义的一个简单的策略是写出研究大纲中所用的概念的清单,然后尝试着用一个词、一种特定的操作(指数计算、观察模式等等)或参考先前研究中所发现的其他概念来定义它们。与同一领域的研究者以及身边的学生反复讨论每一个概念。经常可以发现这样的情况,原以为很简单的概念却包含许多的含义。讨论清楚后才能决定哪一种含义是所想要的。

2. 假设应该可以用经验的事物来表示。前面早已指出,科学的概念必须有一个基本的经验指示物。包含道德判断的假设是无用的,像"罪犯并不比商人更坏"、"妇女应该从事某种职业"以及"资本家剥削工人"这样的陈述,与常见的命题"猪被称作猪是因为它们太脏",或者经典的问题"要为一头黑公牛做出一副肥臀需要多少码的脱脂乳"一样,都是无用的假设。换句话说,当一个假设可能包含对价值判断的研究时,必须将这种目标与道德说教或者接受某种价值的借口分离开来。

该做什么:首先,分析那些表达态度而不是描述或指称经验现象的概念。特别注意那些关键的字眼,比如"应当"、"应该"、"不良的"等等。然后将注意力转到更为有用的概念上。"不良的父母"是一个价值判断的术语,但研究者可以在头脑中有一个

确定的描述:即那些表现出反复无常、专横武断行为的父母,包括对儿童的生理侵犯、缺乏关爱的家长。"应该"也是一个价值判断的术语,但是,学生们可能简单地意指"如果妇女不从事一种职业,我们可以预计,当孩子离开家庭时她们就会发生情感的困难,或者我们可以预计社会将不能生产出这么多的商品"等等。相反,当我们发现我们的指示物仅仅只是一种模糊的感觉,同时,我们不能确定观察它所需要的操作时,我们就应该进一步地研究问题,并发现究竟什么是我们所希望探究的。

3. 假设必须明确。也就是说,所有的操作和这些操作所指示的预测应该十分清楚明白,这样就可以评价实际检验这些假设的可能性。通常,假设是以一种几乎不能检验的一般性术语表达的,其范围也十分宏大。也正是由于它们的宏大,这种重要的思想常常是诱惑人的,因为它们看起来重要且能给人以深刻印象。对于学生们来说,最好避免提出这样的问题,而代之以发展更多的技巧用于更为切实的概念上。

所谓使所有的概念和操作明确化,不仅仅是指概念的明确性,同时也是指对所用的任何指数作出详细的描述。这样,如果假设美国社会中垂直社会流动的程度在增加,就要求指数的应用(在目前,有许多用来确定流动的社会地位水平的操作定义。因此,假设必须包括对其所用的指数的说明。参见第6部分各种合适的指数)。

这种明确的表达方式具有这样一种优点,它可以在我们付出努力之前保证研究的可行性和重要性。同时,它还可以进一步增加结果的效度,因为术语的含义越广泛,就越容易落入采用有选择的证据的陷阱。大多数预言者和算命者的声望依赖于他们陈述预言的能力,所以,几乎任何一种发生的事情,都可以被解释为其预言的实现。我们可以用几乎是统计的术语来对此进行表达:预言越是明确,那么它被证实实际上是一种纯粹意外的结果的可能性就越小。因而,科学的预言或者假设必须通过尽可能明确的定义来避免选择性证据的陷阱。

该做什么:如果一个预言可以分解为更明确的子假设的话,我们就不能满足于一个一般性的预言。例如,关于战争的一般性预言是不够的,我们必须说明时间、地点以及参战者。有关文明全面衰落的预测并不是一个可用于检验理论的假设。同样地,我们必须要能够说明和测量原因、说明下降的意义和时间、说明所包含的不同部分的人口等等。通常,我们可以通过概念分析和相关假设的形成来做到这一点。例如,我们可以预测随着城市化的发展人口生育率会下降。然而,如果我们定义了有关城市化的指数,说明了哪一个部分将会受到影响,以及影响程度如何(因为在美国不同的种族和宗教群体所受到的影响不同),说明了人口生育率下降的总量以及下降的类型(无孩家庭的百分比、净人口再生产率等等),我们的精确性就会增加。我们可以形成子假设:①弄清楚我们所寻求的资料与结论之间的关系;②使特定的研究任务变得更易于处理。

4. 假设应该与合适的技术相联系。在前面,我们就反复指出理论和研究不是对立物,那些不知道何种技术适合用来检验他的假设的理论家,在形成有用的研究问题方面情况也不会好。

但这也并不能作为绝对的教条,用来说明由于当前假设的形成太复杂,以至于做出当代的技术无法对其进行把握的情况,它仅仅只是一种应用于任何问题的早期阶

段,以判断问题可研究性的明智的要求。

不可能的假设也存在一些对假设的形成有价值的方面。如果问题能够重要到足以成为一种可能的参考框架,则无论它是否可以及时地被检验,它都将是有用的。例如,马克思关于社会经济的假设并没有被他的数据所证实,无论在当时还是现在,都缺乏必要的技术。但无论如何,马克思主义的框架却是更为精确的、更细小的、可以证实的命题的一种重要的来源。迪尔凯姆关于自杀研究的许多工作也是如此。他关于社会整合的相关命题也一直是十分有用的。两位巨人的工作对于社会学来说一直是极为重要的,尽管在当时他们宏大的思想还不能用合适的技术来处理。

进一步说,提出不可能的问题也可能刺激技术的增长。现代技术发展的某些推动力显然来自于对某些由于技术限制而被认为是不合适的重要研究的批评。在任何重大的社会学讨论中,研究的边界不断地受到这样一种主张的挑战:即使研究在目前是不可能的,各种问题也都"应该"被研究。

该做什么:查找与研究主题有关的研究论文,列出它们在测量研究中的重要因素时所使用的各种技术。如果你不能找出任何相关技术的讨论,你可能发现基于必需的研究技术来从事这一研究是明智的。或者,你也可能决定,对于你目前的研究资源来说,研究技术的缺乏意味着你的研究问题太大、太一般化。

许多技术一直用于某些项目的研究,比如社会分层或种族态度。尝试着去发现为什么一种技术可以用于一种情形而不能用于另一种情形。注意技术是如何被完善的,看看是否这些技术中有一种对你的目的有用。查找对先前研究的评论,以便于理解后续程序中的弱点。

此外,或许有些问题在没有作出精确测量的前提下被研究了,查阅文献看看为什么这种情况会出现。弄清楚是否某些子领域(例如,宗教行为)受到了运用于其他领域中的技术(例如,态度测量、分层测量、决策研究等等)的影响。

5. 假设应该与一个理论体系相联系。这是一条经常被初学者所忽视的标准。他很可能会选择那些"令人感兴趣"的主题,而没有在意研究能否帮助反驳、证明或支持任何现存的社会关系的理论。然而,一门科学只有建立在现存事实和理论的整体上才具有累积性。如果每一项研究都只是一种孤立的调查,科学就不能发展。

尽管不断增加的理论发展的明显例子的确是在物理学和生物学中被发现,但是,这一过程在社会科学中也可以看到。一组关于智力的社会特征的概括的发展就是一个这样的例子。19 世纪末的人类学研究发现了不同社会中社会习俗的惊人变化,最后显示出社会生活中存在许多共同的元素:家庭制度、宗教仪式、社会化过程中的组织等等。

社会学的法兰西学派,包括 Lucien Levy-Bruhl, Emile Durkheim, Marcel Mauss, Henri Hubert 以及其他人,在世纪之交形成了一系列的命题,认为人类的智力结构由社会的结构所决定。这即是说,感知和思想由社会决定,而不仅仅由我们的眼睛、耳朵以及其他感官的解剖学结构所决定。思想的模式在不同的社会中各不相同。某些表述采用了一种极端的、不关注我们目前状况,同时常常是含糊的措辞形式。不管怎样,一个波利尼西亚土著人的智力不能够用欧洲人的标准来评价。他们的思维不仅具有量的差别,还具有质的不同。

同时，一些更好的技术被用于对“智力”的测量中，最终成为各种智商测验分数的标准化形式。当这些测验被应用于各种不同的群体时，他们在智商方面的变化就变得十分清楚了。意大利移民的孩子在这种测验中的分数较低，黑人的孩子也一样。北方黑人的得分比许多南方州的白人的得分要高。父母为华裔和日裔的美国儿童的得分更高。由于通常假定这些测验所测量的是“天生的智力”，因此这些数据有时被用来得出某一特定的“种族”群体先天就是劣等的，而另一些则是优等的结论。

然而，这种结论与理性的依据相反。开明的观点认为，这些结论应该接受检验。于是就有了两组主要的结论：一种结论认为，智力主要由社会经历所决定；另一种结论则认为，智商是先天决定的。为了检验这样两种相反的概括，就需要一种在更为特殊的情形中来检验逻辑结果的研究设计。例如，如果“劣等的”群体中的成员的智力真的是由于生物因素所决定的，那么，他们生活环境的改变将不会改变他们的智商。另一方面，如果社会生活经历是更基本的因素，我们将期待这样的结果，即当人们的社会生活经历发生改变时，他们的智商也会发生一定形式的改变。

我们还可以做进一步的推演。如果将双胞胎分开，并在早期将他们分别置于完全不同的社会生活经历中。那么，我们可以期待他们的智商将会有着显著的差别。或者，如果一群农村的黑人儿童从南部贫穷的学校和社会生活经历中迁移到北部多少有些刺激的环境中，群体的智商平均分也应该多少有些改变。在一项经典的研究中，Otto Klineberg 进行了后一种探索。他跟踪研究了迁移到北部的各种年龄的黑人儿童，他发现，总体上，越早搬迁到北方，智商平均分的提高越大。搬迁越晚，智商平均分增加得就越少。即使我们假定那些比较好的、比较能干的、比较大胆的成年黑人也进行了这样的搬迁，它也不能解释搬迁时间所带来的差别。此外，当然还有在搬迁时对象是孩子的因素。

在这一研究设计中，由一大组概括所得出的一系列演绎预示了一种特定的结果。进一步说，预言实际上得到了证实。要对那些一直从事于修订和发展智商测验的大量学者做出公正的评价，还应该提到许多人类学家、社会学家以及社会心理学家也一直在进行着的其他类似的测验和研究。他们并不否定智商部分地基于“天生的”能力的看法，但他们明确指出，在很大程度上，这种能力必须被某种类型的经历所激发才能够在这种测验中获得高分。

即便是从前述的理论发展的初略轮廓中，我们也能看出，当研究越是系统地基于某种现存的理论体系时，就越有可能在知识上产生出真正的贡献。换句话说，要使研究有价值，假设不仅必须仔细地陈述，它还必须具有理论相关性。

该做什么：首先，当然是要查阅与你的研究主题相关的文献。如果不能这样做，那么你的假设就可能会涉及太宽泛的领域。第二，从文献中抽象出各种命题和命题组之间的相互联系的方式（例如，与犯罪学中 Sutherland 的差别联合理论，企业中增强士气的条件，或者婚姻调适的预期研究相联系的文献）。第三，一个一个地，或者从一小组主要的陈述开始，确定是否可以减少你的命题，包括你自己的假设。第四，用某些理论模型，比如 Merton 的“社会学中功能分析的范式”（*Social Theory and Social Structure*, pp. 50-55）来检验它，看看你是否丢掉了一些重要的命题和决定因素。第五，特别地，

将你自己的相关命题组与那些经典作者的进行比较,比如韦伯关于科层制或者迪尔凯姆关于自杀的命题。如果你发现这种摘要的任务很困难,那么也可以换之以比较系统的解说者,比如 Talcott Parsons 在其《社会行动的结构》(*Structure of Social Action*)一书中所解释的那些人的命题。最重要的是,无论你的假设的来源是什么,它都必须能够逻辑地从一组相关的社会学命题中导出,并且必须以一组相关的社会学命题为基础。

注 释

1. Otto Klineberg, *Negro Intelligence and selective Migration*(纽约:哥伦比亚大学出版社,1935)

资料来源:William J. Goode and Paul K. Hatt(1962),Methods in Social Research,(纽约:McGraw-Hill),pp. 68-73

关于理论和研究的推荐读物

Fawcett, Jacqueline. (1999). *Relationship of theory and research.* Philadelphia: F. A. Davis.

Gelso, Charles J., & Hayes, Jeffrey A. (1998). *The psychotherapy relationship: Theory, research, and practice.* New York: John Wiley & Sons.

Mongardini, Carlo, & Tabboni, Simonetta. (Eds.). *Robert K. Merton and contemporary sociology.* Unpublished manuscript, Department of Political Studies, University Roma "La Sapienza."

2.10 科学:即时观察的、实验的和历史的

对于处于事业起始阶段的科学家来说,决定采用何种方法去研究一个其所感兴趣的特殊问题一直是个不小的挑战。考虑是用历史性的分析、统计性的采样、定性的结构观察、受控制的实验,还是需要其他的方法这一过程是十分困难的,并且在很大程度上取决于研究所涉及的研究者自身的经验。

在接下来的内容中,Raymond Siever,一个物理学家,描述了科学的多样性和不同的类型,并强调了问题的重要性及其与科学方法的关系。

科学:观察的、实验的、历史的

Raymond Siever

"到底是有很多种科学呢,还是只有一种?",这个问题自从科学开始互相区分就一直困扰着许多科学家。由此就又引发出另一个与此相关的问题,即"只有一种科学方法还是有很多种?"。讨论这些观点的时候,常常会受到发言者的个人背景,尤其是他手头上正在做的科学研究的影响。同时,也会由于混杂的主题而与研究如何操作执行这类问题相混淆。我将会对科学传统上的各种分类之间的相关提出自己的看法,并

且对是否存在一种或多种科学提出一些探讨性的答案。这并不是因为这些观点是新的，而更大程度上是我们需要提醒我们不要忘记哲学的基础，特别是当今科学的分支已经越来越细化，而且当它们同时又在复杂的系统中相互融合互相攻击的时候，更是如此。

观察的与实验的

观察性的科学与实验性的科学之间的差异是十分明显的。在本文的有些措辞中，“观察的”一词与“单纯性的描述”的思想是相关联的，而“实验的”一词则常常与分析的方法相联系。因此，一些视自己高人一等的科学家就将这些进一步扩展，暗示科学存在好坏之分，他们将观察性的科学看作是坏科学，将实验性的科学当作好科学。对于术语的这种选择性已经在科学界成为不成文的规定，因为将在另一个领域的同事的工作称作差的，这不太明智，而将其简单的称为“描述性的”将会更好。我们都知道，对于好的和坏的这两个词都有一个恰当使用的问题，可能仅仅将其应用到个人的工作上时才是合适的。

当然还有许多我们熟悉的术语。比如有“硬”科学，我猜测这也暗示着有“软”科学。我们也知道另外还有许多的词语可以将科学中的“两种文化”区分开。在试图不破坏规则的情况下，值得指出这些词语(如观察性的、描述性的、实验性的、分析性的等)是如何使用的。

被用来描述科学之间差异的词语	
分析的	描述的
实验的	观察的
软的	硬的
非数学的	数学的
好的	坏的
有趣的	枯燥的
合趣的收集	枯燥的实验
古典的	现代的
普适方程	百科全书式的专论
严格的	不精确的
简单的	复杂的
批判性的	墨守成规的

我认为，所有的科学都既是观察的又是描述的，这一点是毫无疑问的。一个例子就是，伴随了我们很久的热流方程式，它完全是在基础性的观察上得到的。牛顿最初对于热流所建立的规则比现在我们所使用的更为精致的数学陈述要简单得多。但是这个使我们能够做很多事情的精巧的公式表达是建立在相当基础的观察之上的。所以谈论什么非观察性的科学或者是非描述性的科学不是一件明智的事情。

据说有一些科学家仅仅是描述事物而不对它们进行分析。他们说为了事物本身而进行的描述就值得称作是科学了。当然，许多科学在它们的初级阶段是以大量的数

据采集而不是数据分析为特色的。这不是暗指那些采集数据的就没有想到过他们在描述什么或是试图将它们归于某种模式中去。很明显,那些描述的人对于他们将要描述什么事物是做了决定的,而且他们的分析过程就在于对于这些事物的选择之中了,或者是,虽然这个观点仍需探讨,或许就是第一个人登上月球将会在什么时间。

无可否认,从事科学研究的人中的确包括了一部分为了自己的原因而进行描述的,他们承认描述是他们惟一的目标。他们与科学之间的关系就如同收银员与生意之间的关系一般。但是大多数单纯描述的人会说他们这么做只是暂时性的,因为他们始终还是为了分析这一目标而努力的(时常是推延到未来的某个时间)。

如果的确是仅仅为了描述而描述,而不加以说明描述的是什么或是如何在描述之后将它们整合起来的话,那就不是我们通常所说的科学,那我们就真的不能谈论什么描述的或非描述的科学。当一些科学家说起其他科学家的工作或者其他科学领域的时候,他们会说"那是描述性的",其实他们真正想说的是那算不上科学。

说得好听点,以上的论述也可以被解释成,与其他领域相比,在这些领域中的描述性所占的比例比分析性所占的比例要高。当然,它们的比例随着各领域的发展阶段的变化而变化,随着研究者的不同也有所差异。即使在一个已经大大超越了描述多于分析这一阶段的领域,一个新仪器的发明也会导致新的观察,同时也会产生出大量的与分析相关的数据。

如果科学努力的目标之一是从某一特殊的观察中抽出普适性的规律的话,那么必须承认,这一努力或多或少有些艰难。物理学基本上已经变成了某些人的领地了,这些人可以从有限的数据中相对简单地做出概括(虽然还没有人声称物理学是一项简单的智力劳动)。另外一种解释的方法是数据之间变化很小,概括得很好。在一些特定的领域中,也许社会科学是最明显的例子,数据之间的变化太大,以至于很难对它们进行归纳。这就不可避免地导致各学科整体逻辑结构的不同。物理学的一大部分是与可以推导出所有其他部分的基础物理学理论紧密相连的,即所谓的第一定律。另外,在一些领域,例如工程学的某些领域,经验主义就是其指导规则,这只是因为那里不存在普遍有效的"第一定律"来指导具体的操作。

实验与科学

实验总是与科学相连的,而且被看作是科学的一个最强大的工具。我们对于实验的看法大部分是基于物理学与化学中所做的工作。但是其实我们可以从很多角度来看实验。它们可以被分为控制实验与非控制实验。从选择的角度看,我们可以将实验表达为自然的或是人为的。人为的实验是指,选择好反应材料和实验所需的环境,然后就观察实验活动的过程或是最后的结果。

除了那些主要的兴趣在于生物学、地球科学或是天文学的人,我们也许对于自然的实验不怎么熟悉。我们也许会问,如果牛顿那天是在地上看到那个神奇的、熟过头的、被啃掉一部分的、已经腐烂的苹果的话,将会发生什么事情呢?从这样的观察中,他还能抽象出关于地球重力的一般法则来吗?在我看来,可能性不是没有的,但是也许时间和努力的耗费上就会更多,而且具有更大的不确定性。例如,许多地球化学家

用某一特定的方法考察地球上的化学进程。这就好比是一个为了弄清在基础化学实验室中发生了什么的人,进入到那个没有人在的实验室中,分析他在水槽中和在实验室伸出的下水道中所发现的东西,记录下实验室是如何配备的,使他有可能推论出刚刚所做的实验,并且猜测所采用的是何种试剂。当然,自然的实验也和人为的实验一样具有局限性:观察者必须选择正确的观察参数。

自然的实验可以通过观察其不同的部分,或是选择那些特别控制或限制的实验所得到的偶然事件而加以净化。在多变量的情形中,我们寻找那些变量少一些的时间或场合。那些花费了大量的时间追寻控制性自然实验的人,会深有感受地诉说,找到一个完美控制的自然实验几乎是不可能的,自然实验总存在缺陷。而且那些利用这些数据工作的人似乎总是试图从这些差劲的实验中推导出概括性的内容。

科学中,对于人为实验的可能性的限制是很多的。第一个限制是某些系统过于庞大,换算系数并不是每次都能得到,而且也不足以将系统减小成实验室中的检查所需的大小。最著名的两个例子就是天文学和地理学。同样,尽管庞大系统的受限制的零碎部分可以被移动并带进实验室中,但是系统之间的相互关联性本身却不能够被复制。

一些系统的复杂性和内在联系性限制了实验。Warren Weaver(1995)将"强结合"与"弱结合"应用于科学中。Weaver 运用这些词语来将自然科学从社会科学中区分出来。但是我认为这一想法也可以被用于自然科学内部的区分。例如,对于海洋研究的某些部分,总体的海洋循环表面看上去也许是弱结合的,因为人们很少会想到在地球的移动、大气层以及地球和海洋的热量收支这几个变量之间会有什么内在联系。海洋学的另外一支——生态学,则是一个强结合的学科。因为海洋生态联系紧密,以至于很难将各个变量分开。似乎地球表面绝大部分的自然现象都是紧密相连的,不管是为了实验还是分析的目的,都无法将它们分开。

当然,现在有大量的人为实验,并且已经揭示了很多的信息。我就会将现代的空气和水体污染这一人类的灾难归类为一个典型的实验,虽然它给社会带来了不幸。我可以提供更多的例子:C^{14}这个炸弹在大气中传播,并和海洋发生了反应,这给我们呈现了一幅更好的 CO_2 循环图,并且在海洋和大气之间达到了前所未有的平衡。企图阻止五大湖现有的污染可以看作是一个相反的实验:我们对于五大湖是否可以被清理干净,虽然充满了希望但是从科学上说又有着不确定性。不管结果如何,污染及其对策教给了我们许多的科学(或是工程学的)知识。

过去,社会禁忌阻止了许多实验的进行,但是现在,即使是这些社会禁忌有时也都会被打破,最著名的就是在某些纳粹集中营进行的所谓"实验"。已经有人这么说道,越南战争就卷入了一些新武器装备和新观点的实验性测试。但是有一点在大多数时候都是正确的,即对于科学家来说,生物学实验中被认为是重要的领域,一旦考虑到足够多的社会原因的时候就成为了禁地。

模拟或假设的实验和系统分析被用来代替社会控制或是那些过大以至于不能带入实验室进行的系统。但是这样的实验也不过是与可以在头脑里进行的第一定律差不多。理论物理对于具有如此价值的实验是一个很好的选择。但是在世界上大部分科学实践中,科学家运用假设性实验作为真正实验或是进一步观察的序曲。在这一过

程中,科学家不是从他的兴趣出发进行假设性实验。科学家进行假设性实验应该就像向自己的老师求教一样,是非常自然的事。在假设性实验中,实验人员自身的作用是微乎其微的。

看来,尽管实验科学种类各异,尽管看的角度各异,但实验的本质都是一样的,控制的手段可能会有不同,观察实验的不同部分的能力可能有限,最终还是有些仅仅由于社会原因不能做的实验。

历史性的与非历史性的科学

这个在科学本性的对话中不断重复的主题(Nagel,1952),最近又被 G. G. Simpson (1963)和 R. A. Watson(1966)重新探究。在我看来,除了在经济上可能更有利可图,或者在文化上需要精确地确定在某一特定的时间和地点到底是发生了什么以外,似乎在历史性和非历史性的科学之间没有什么本质的区别。因此,我们实在不太关注如 Watson 提到的,科罗拉多河的大峡谷是如何形成的之类的问题。假设峡谷的分割不是一个特殊的事件的话,我们只是关心一般性的大峡谷过去是如何形成的。这就好比一个化学家是不会去关心在某一个实验中的具体数据,他惟一所关心的是,重复和概括实验,从而使得他和别人的相同实验的结果会落在差不多的模式之中。实际上,一个人很少关注实验中的某一特殊数字。人们对于原始数据并不真正关注,除非它们是作为计算那些人们真正关注的数量的媒介,而这些数量的重要性在于使以往的科学研究建立起来。所以,虽然我们测量了特定的质量和体积,但是我们只从中引用密度这一重要的数据。

我们可能会用进程中的时间刻度来将历史性的科学与所谓的非历史性的科学区分开来。虽然一个化学反应也有一段"历史",但是这个"历史"比我们所认为的"历史"进程要短得多。即使是很慢的化学反应也比地理进程要快得多。在天文学上也一样,很多过程都是相当缓慢的,虽然有些其他的过程也许会快一些。但是即使是化学反应的历史也是十分重要的,因为化学动力学研究的就是这个。此外,虽然这是一个历史的事件,但是研究反应进程的化学家更感兴趣的是那些任何人都可以做的可重复的实验,而不是那些在他的实验室中某一时间进行的某一反应过程的化学研究。

对于历史性科学不同之处在于,许多时候只有一个自然实验是可观察的,以至于使概括变得非常困难甚至不可能。在地球上,显然我们只有一个机体宏观进化的例子;在地球原始的大气中,氧气的全面出现大概只会发生一次;在现代社会中,由原子弹所引起的我们生活的变化可能只会发生一次。如果实验的本质——不管它是人为的或是自然的——在于它是可重复的,并且人们需要起码的自由来对此做出平均或是概括,那么,我们对于某些具有惟一性的事件就束手无策了。这不等于是说,它们在宇宙中就是惟一的,只是在我们的观察能力所及的范围之内它们才是独一无二的。正因为此,在生物学中才会存在着对于在月球或是火星上是否存在生命形式的可能性的兴趣,他们只是在寻求额外的实验。比独一无二的实验更糟糕的是获得那些变数很大而又为数不多的实验:在地球上,我们只有几个大洲;在北美洲的结构发展中,只有很少的几次地槽的和山脉链进化的主要进化模式;只有很少的陆地行星。社会科学在某种

程度上受着同样的折磨:至今只有几个国家拥有了原子弹。

科学的风格

每个科学家在选择自己的学科时都有很多的理由,但是有些风格在所有的学科中却是共同的。我在这里使用了"风格"一词是因为,我不认为存在不同种类的科学,或是科学学科。正如Kurie(1953)和Eiduson(1962)指出的一样,在科学中渗透进了许多不同的个性,而且每一种个性类型在处理事情上都有自己不同的方式。虽然在个性与所选择的学科之间存在着一些相关,但是我并不在此阐述这些。

风格是一个有很多种含义的词语,从学科中的一个特别的历史流派(如经典风格)到一项特别的方法名称,再到特别的时代、研究领域与科学家独特的个性交互作用而产生的智力结晶。在这里,我所使用的风格仅仅是指后一种意思。风格可能与个性有关,但是它们也总是受到个人所工作领域的限制。最近关于不同风格的典型的例子就是两个诺贝尔物理学奖得主Richard Feynman(1966)和Julian Schwinger(1966)的演讲致辞的对比。这两个在物理学的同一领域工作的人在解决同类问题以及描述它们的时候表现出了截然不同的风格。

我们可以辨别出那些更为与众不同的风格并给它们贴上标签,而这些风格在所有领域中都是很普遍的。我们会发现它们中的一些是相互交叉的,甚至有的可以在不同的时间里或是在状态冲突的时候变换几种不同的风格:

- 严格的形式主义者
- 聪明的现象学家
- 勤恳的实验室方法学家及与其等价的认真细致的田野工作者
- 麻利且龌龊的撇乳工
- 瞄准机会的情人或是地位相等的垄断者
- 准通才或是地位悬殊的垄断者
- 按成功和失败被划为业余艺术爱好者或多才多艺的名家
- 睿智的长者

这是一个室内游戏,每个人都可以和自己的朋友和同事尝试一下。

人们常常对不同流派贡献的相近价值进行价值判断。但是可能由于所有这些风格对于科学的进步都是必须的,每个人都依赖于其他人。现在为挑选科学中的领导,而对某些风格给予过多的关注,这种现象是危险的。这种风格又将会与学科以及推进科学前进者的个人的基本能力相混淆。多元性与多样性有助于产生更多的科学兴趣,正如我们在生活中的其他领域所见一样。但是让我们在风格和主题方面享有不同,并且让我们认识到不同种类的科学之间不公平的差别仅仅是为了建立地位和特权的等级制度。

资料来源:Raymond Siever(1968)."Science: Obserrational, Experimental, Historical," *American Scientists*,56,70-77.

2.11 设计、模型建构及大规模研究指南

以下内容是强调建立大规模经验研究实践的读物。开始的一篇是由 Samuel A. Stouffer 撰写的"对研究设计的某些观察",被认为是关于大规模社会研究设计要求的最有用的单篇论述之一。

模型建构已经成为科学工作的一个完整的部分。"模型在研究设计中的角色"(2.12 小节)中,描述了当前使用的各种类型的模型。

Edward A. Suchman 在"研究设计的一般考虑"(2.13 小节)中,列出了在理想的计划必须妥协时所经常需要的一些现实的评价。

2.14 中所描述的是影响研究设计的内在效度与外在效度的因素,它包括 Donald T Cambell 和 Julian C. Stanley 关于实验设计专论的一部分。大规模的群体研究在数量上和在范围上都已得到发展。

最后,2.15 小节回顾了抽样的各种类型,包括抽样类型、简要描述及每一种抽样类型的优点和缺点等。

对研究设计的某些观察

Samuel A.Stouffer

我们必须清楚证明是由什么构成的。如果可能,我们必须提供生动的例子,而不是似是而非的优点,说明信赖某些东西。问题的核心位于事先的研究设计中,因此,证据不能只是一打可以替换的解释。

基本上,我认为最根本的是我们要将控制实验的模型始终记在心中,即使在实践中我们可能不得不背离理想的模型。见下列简单的图表。

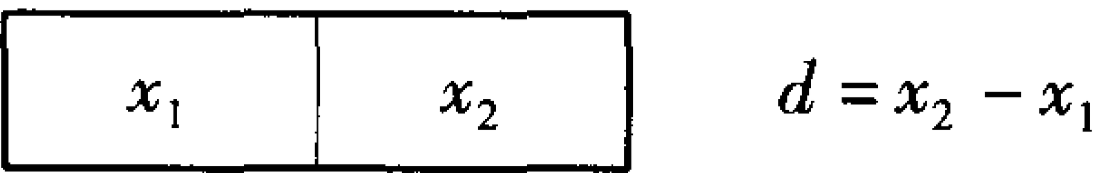
	前测	后测	前后测之差
实验组	x_1	x_2	$d = x_2 - x_1$
控制组	x_1'	x_2'	$d' = x_2' - x_1'$

要检验差 d 是否归因于我们所认为的原因,就要看 d 是否显著地大于 d'。

在战争期间,我们反复地应用这一模型去测量带有某种取向的电影对士兵态度改变的影响。如果我们不采用完备的实验设计,只是简单地用一个随意选择的男性样本,来对那些说他们已经看过这部电影的和那些说他们没有看过这部电影的人进行比较,那么,我们会得到印象相当深刻的差别。这对我们也是一种很好的回报,因为管理者希望电影是强有力的药品。这种欺骗性是由于记忆选择的不可靠性。那些清楚记

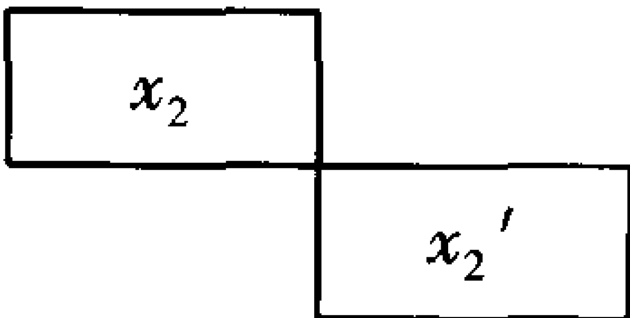
得所看电影的可能是那些对信息最敏感的人,而那些无兴趣的或漠不关心的人实际上可能看过但自始至终都在睡觉或者仅仅是因为忘记了。

大多数情况下,我们不可能或者没有足够耐心以保持上述图表中所有四个格子的方式来进行研究设计。有时我们仅仅只有顶端的两个格子,如下图所示:



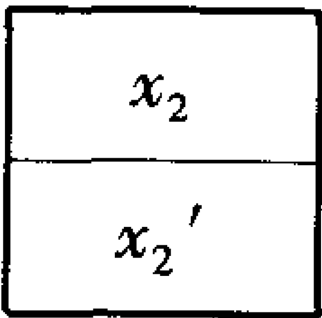
在这种情况下,我们具有对同一个体或群体在不同时间的两次观察。这经常是一种非常有用的设计。例如,在军队中,我们可以调查一组新兵的态度,以后再次研究他们。从中我们可以了解哪些人的态度发生了改变,以及朝哪个方向改变(其结果几乎总是朝着更坏的方向变化,这就使得我们总是不被军队所欢迎)。但是,对于在最初的训练期中,究竟是哪些确凿的因素最应该对态度的变坏负责,却只能靠间接推断。

同期群研究通常比常见的设计得到的信息更多,它可以用图形表示如下:



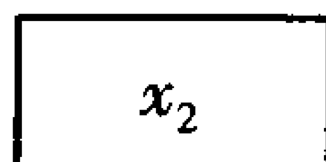
这里,在时间的一个点上我们有一个样本,在时间的后一个点上,我们有另一个样本。我们观察到,我们的测量,比如说,均值,在较近的样本中比在较早些的样本中更大一些。但是,我们从观察中不能确定哪些人或者哪种类型的人发生了改变。同时,总是存在着这样一种令人烦恼的可能性,即我们两个样本所取自的总体一开始就是不同的。因而,差别或许不能归因于发生在两次观察之间的这段时间内的条件变化。这样,我们研究在美国本土的一群士兵,后来又向海外的一群士兵询问同样的问题,通过对这两个群体的人在服役的部门、从军时间的长短、军阶等等方面进行仔细的配对,我们希望研究的结果与对同样的人研究两次所得到的结果相近似。但这或许只是一种希望。某些重要的因素,比如身体条件,不可能充分地被控制。那些到海外的人通常一开始就比那些留在本土的人有更好的体质。然而,如果后续的研究在热带地区进行,那么,不利的气候就会使他们不得不为此付出代价。有多少海外的士兵发生了改变?如果我们希望对研究的发现有更多的信心,就需要把进行同期群研究作为最低限度。

一个常见的既不付出代价但又得到控制实验结果的办法是采用如下图所示的设计。这就是通常我们用相关分析所得到的。我们有两个或多个群体,我们在同一时间点上研究它们。

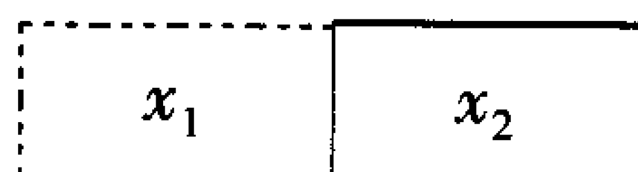


这样，我们有在步兵的人、有在空军的人，我们比较他们的态度。但是，我们不能确切地知道，x_2'与x_2之间有多少差别可以归因于在某种服役部门的经历，有多少差别是受每一个部门所选择的人的分布的影响。确实，我们可以通过配对来排除各种可能性。例如，我们可以比较来自两种不同部门，但具有相同年龄、相同文化程度的人。但是，其他一些未被控制的变量可以通过一扇经常敞开着的宽宽的大门进入。

无论你是否相信，有时我们只有一个格子：



当这种情况发生时，我们对一切都了解不多。但是，如果我们从我们的想像力出发，用似是而非的推测代替缺失的格子，我们仍然能够用“有才气的分析”来填满社会科学杂志的页面。比如，我们可能发现今天的青少年具有疯狂的思想，于是得出社会正在走向堕落的结论。我们在下图点划线格子中填上象征我们自己昨天的假设的资料，x_1象征我们自己，x_2象征我们的子女。令人哭笑不得的是，大多数公众，我担心，包括许多的社会科学家，他们对文化已经如此适应以至于到了无需追问原由的地步。



我并不想贬低所有与控制实验的规则不一致的研究。我想我们在社会学和社会心理学中将会看到比过去更多的完全的实验设计。但我也意识到它在执行中会有实践上的困难，同时我也了解存在着无数的重要情形，在那些情形中，它一点都不切实可行。我所强调的只是对缺少了那些关键性格子的设计的局限性的认识。

有时，如果我们小心地避免引申过甚的解释，我们可以通过预先设计和事后拼凑得到有用的相似性。让我引用一个例子：

在战争时期的欧洲，军队检验了那种把一整个排的黑人士兵放到白人步兵团中的想法。在几个连中这样做了，黑人士兵与白人士兵并肩战斗。几个月以后，我们被邀请去了解白人军队怎么看这项改革。我们发现，在与黑人士兵排一起打仗的这几个连中，只有7%的白人士兵说非常不喜欢这种想法，而在那些没有与黑人士兵一起打仗的部队中，则有62%的白人士兵说，如果在他们的部队中做这样安排，他们将非常不喜欢。我们有如下结果：

	7%
	62%

现在的问题是，这些与黑人士兵并肩打仗的白人士兵，是否原本就比那些没有与黑人士兵并肩打仗的白人士兵更赞成与黑人在一起打仗呢？我们不这么认为。因为，

比如说,他们包含同样比例的南方人。然而,如果我们打算作出"与黑人士兵在一起的实际经历使得敌意从62%减少到了7%"的推断,这一点就十分重要。作为一种仅次于最好的选择,我们在那些与黑人士兵并肩作战的连队中询问白人士兵,他们是否能回忆起当这种改革第一次被提出时他们的感觉。结果,67%的人说他们最初是反对这种想法的。这样,我们就能够暂时地填上缺失的空格,并得出结论说,在所具有的条件下,白人士兵的态度有了一个明显的改变。

即使这是一项完全的控制实验,也存在着大量的得出错误推断的可能性。研究的结论仅仅只能应用于那些与这一研究非常相似的情形。比如,有时候所包含的黑人是那些志愿离开后方地区去战斗的人。如果其他的黑人也被包含进来,那么情况就会有所不同。此外,他们还有白人军官。军队中一位看过这一报告的陆军少校对我的祝贺令我十分吃惊,因为他过去反对改革,我原以为他会奚落这项研究结果。他说,"这项研究结果证明了我在三十年军队生涯中的一贯主张——只要给黑人士兵配备白人军官,他们一样会干得很好!"同时,这项研究结论仅仅只能应用于战斗经历中。要证明将它推广到非战斗的或卫戍部队的情况下也成立,则还需要其他的一些研究。换句话说,一项单独的研究,无论其设计得多么好,如果超出了它直接的关联来使用它,就可能是一件非常危险的事情。

目前,实验要花费时间和经费,认为我们在社会科学中不能像生物学家那样,通过在各种土壤和条件下培植庄稼,同时地开展成百项实验,是没有用的。实验在多数——并非全部——自然科学中是相对容易的事实,成功地说明了物理学家、生物学家作为一方,而社会科学家作为另一方的两方在所要求的证明质量上的差别。

尽管当我们需要的时候并不是总能设计出完美的实验,但我们却可以至少在我们的眼前保持实验的模型,并且在填充上述点划线的缺失格子时谨慎从事。然而,考虑到经济方面的利益,我们可以采取进一步的,同时也是更为重要的行动,而这就要依赖于我们对初始问题的选择。

参考文献

Hovland, Card I., Lamsdaine, Arthur, & Sheffield, Fred D. (1949). Experiment on mass communication. Princeton, NJ: Princeton University Press.

资料来源:Samuel A. Stouffer (1950). "Some obervations on study Design," American Journal of Sociology, 55, 356-259.

模型在研究设计中的角色 2.12

建构模型是社会科学整体的一部分,因为模型指导着理论发展与研究设计。在当

代生活中,建构模型已受到重视,并被许多力量所推动。在计算机、生物工艺学和自动化的世界中,模型都是十分适用的,而且它们还使政府的、工业的和军队中的科学家有了新的地位。对于社会科学家而言,模型也同样重要,因为它们在探讨重要问题的时候提供了框架。

为了进一步强调建构模型这一活动在社会科学中的重要性,“建模”(model building)已经拥有了代表一项行为的动词的地位。尽管我们仍然使用“模型建构”,但是这一活动已经越来越多地被“建模”所代替,而现在社会科学的语言中包括了这样的术语,如游戏模型(game models)、模拟模型(simulatioin models)、数学模型(mathematical models)、趋势模型(trend models)、随机性模型(stochastic models)、实验室模型(laboratory models)、信息与控制论模型(information and cybernetic models)、因径模型(causual and path models)等等。即使是理论自身也被分成了好几个理论模型。

所有这些术语都代表了一个产生预言(或是假设)的封闭系统,这些预言或是假设产生以后需要一些经验的测试。应该鼓励研究者运用模型,帮助确认重要的变量,以便可以更加准确地对假设的检验进行阐释。

在试图给各种模型进行某种排序时,我们很容易发现有大量的重复之处,并且也存在广泛不同的用法。一直以来,对于模型分类的看法就没有完全一致过,但在以下的描述中,我们列出了五类模型以及它们的变量。

- 物理(实体)模型(physical models)
- 理论模型(theritical models)
- 数学模型
- 机械模型(mechanical models)
- 符号互动模型(symbolic interationist models)

只要条件允许,研究者就应当以如下的规则来指导模型的设计与使用:

- 我们首先必须透彻地理解要被建模的系统。
- 只有一个系统中的重要部分,以及它们的控制量才可以被建模。
- 有可能的话,模型中的常数、比率和相关系数必须通过实证性的测量得到,而不是直接从文献资料中获取。
- 建模的练习以及相应结果的模拟可以看作是深化我们理解系统如何运作的工具。

参考文献

Hobbie, James E. (Ed.). (1980). *Lymnology of tundra ponds, Barrow, Alaska.* Stroudsburg, PA: Hutchinson & Ross.

Lave, Clarles A., & March, James G. (1975). *An introduction to models in the social sciences.* New York: Harper & Row.

2.12.1 物理(实体)模型

物理(实体)模型是一个能表现出如同模型使用时现象的具体有形的物体。这些

物体具体表现出静态的或是结构的特性,例如,骨骼、器官、分子、原子、微缩的大楼、飞机、风洞等都可以用物理(实体)模型来演示。可能在当代科学中最著名的模型要算是显示控制人类遗传的 DNA 密码基因结构的双螺旋结构模型了。在许多领域内,领示操作模型(pilot operating models)引入动力系统模式来象征运行着的机械装置。

从缝纫到建筑和航空工程学,物理(实体)模型在几乎所有科学的领域和科技的部门中,都发挥了促进认知的功能。在社会学中对于物理(实体)模型的运用有限,但是 F. S. Chapin 在实验中运用模型来演示公共机构和社会空间,D. C. Miller 也用模型来研究团体与权利之间的关系。

关于物理模型的推荐读物

O'Meara, Tim. (1997). Causation and the Struggle for a Science of Culture. Current Anthropology, 38,399-418.

2.12.2 理论模型

模型这一术语常被用来泛指任何的以短语出现的符号的或是形式的科学理论。如果认为将理论与模型作为同义词使用有什么价值的话(也许许多社会科学家都不会同意),那么它可能就在于:当一个理论被作为一个假设提出的时候,其内部之间的关系能够被清晰地说明和展示。

关于理论模型的推荐读物

Dubin, Robert. (1969). *Theory buiding: A practical guide to the construction and testing of theoretical models.* New York: Free Press.

Land, Kenneth C. (1971). Formal theory. In Herbert L. Costner (Ed.), *Sociological methodology* 1971. San Francisco: Jossey-Bass.

Lave, Charles A., & March, James G. (1975). *An introduction to models in the social sciences.* New York: Harper & Row.

2.12.3 数学模型

数学模型一词,当被应用于社会科学中时,是指运用数学方程式来描述人、群体、团体、州或国家的行为。在趋势模型、因径模型、随机性模型和游戏模型中我们可以观察到对于数学模型的普通运用。

例如,一种数学模型——趋势模型,是关于时间序列对于方程式或曲线的适应性,被假定为原则或法律的替换。因径模型则包括了建构一个社会现实的简单化的模型,其中的变量被假定是遵循因果关系次序的。最重要的影响因变量或评判标准的变量是根据它们的影响力或效果来寻找和安排的。所有其他进入因果系统的变量都被认为是残差。

随机性模型是指一种概率建构,在这种建构中,一系列的行为事件及时发生,同时

被赋予其中的还有这些事件联合发生的概率。这些模型都是处理“随机过程”的。

游戏模型是建立于数学理论之上的，是指在有两个或两个以上的个体或团体参加的竞争性的环境中(策略游戏)，对于最佳策略的确定。和结果完全靠运气的游戏相反，策略游戏的结果是依靠，或是完全取决于游戏者个体所选择的步骤。

关于数学模型的推荐读物

Backman, Olof,& Edling, Christofer. (1999). Mathematics matters: On the absence of mathematical models in quantitative sociology. *Acta Sociologica*, 42,69-78.

Manfredi, Piero. (1994). Mathematical models for communication: An introduction. *European Journal of Population*, 10, 384-386.

2.12.4 机械模型

在社会科学中，机械模型运用来自于用物理学和工程学的概念对社会行为进行类推。随着人们对于数学模型日益关注，对于这些机械模型的兴趣也日益增长，因为机械模型是建立于数学语言和符号逻辑的基础之上的。普通的计算机模型是机械模型的中心，也是使用计算机模拟模型和电子模拟模型的术语。

计算机模拟模型专注于使用计算机程序来提供对于一系列结构体的测试，这些结构体内部相互协调，并具有假定的从经验型数据中提取出可概括特性的解释性能力。模型的假设可以编入电脑程序(让电脑将理论变成程序)，然后由电脑执行程序(理论)所规定的行为。

微量分析模拟模型(microanalytic simulation models)被用来检验不同类型政策对于人口的统计学结构、储蓄与纳税行为、退休后的收入安全、健康与疾病的社会阶层差异，以及人口与其福利的其他众多方面的影响。微量分析模拟系统(MASS)最初是由耶鲁大学的 Guy Orcutt 发展起来的。它是一种运用电脑来获取国家的社会和经济结构的方法。它足够精密复杂，可以应付复杂如真实生活事件般又对整个人类的经济生活有影响的研究问题，如结婚和离婚、生小孩、变换住所、失业和重新就业、退休和死亡。能够包含有普遍的模拟模型来获取如此大量的人类行为的研究仍很罕见。

关于机械模型的推荐读物

Panel to Evaluate Microsimulation Models for Social Welfare Programs. (1991). *Improving information for social policy decisions: The uses of microsimulation modeling* (2 vols. , edited by Constance F. Citro & Eric A. Hanushek). Washington, DC: National Academy Press.

2.12.5 符号互动模型

符号互动模型是针对行动者给予他们所运用和所遇到的符号的含义而言的。在社会互动中，行动的暗示是通过语言和手势来传达的，在交易进行中，行为也是不断发

生变化的。一些符号互动模型是简单的仅包含几个人的结构体;其他的则较为精细,且使用计算机来寻求范式和普遍性。所有的模型都往往是模拟模型,也就是说,它们是基于和现实情景同构的人为情景或是结构概念的。

实验室模型(符号互动模型的一种类型)包括了根据研究者的设计,模拟的演员扮演结构性或是非结构性角色的团体或组织的人为情景。总地说来,这样的行为是在一个拥有观察和记录仪器的封闭的环境中被观察的。小团体实验室中可以引证一些著名的例子。过去,实验室模型受到的关注较少,目前却发展迅速。

信息和控制论模型描述了在交流系统中的信息输入、流程和结果。这些模型的范围可能包括从机械性到符号互动模型,它们所具有的意义也更重要。最普遍的相似物是人类的智能,我们也注意到了运用超级计算机对这一特征进行模仿的增加。

关于符号互动模型的推荐读物

Manis, Jerome G. , & Meltzer, Bernard N. (Eds.). (1967). *Symbolic interaction: A reader in social psychology*. Boston: Allyn and Bacon.

Schneider, Anne L. , Snyder Joy, Zoann, & Hopper, Melanie. (1993). Rational and symbolic models of attitudes toward AIDS policy. *Social Science Quarterly*, 74, 349-366.

研究设计的一般考虑 2.13

下面,Edward A. Suchman 描述了在科学的过程中,研究设计的角色是什么,同时强调了有好几种方式来做"正确的"事。

研究设计的原理

Edward A. Suchman

1. 对我们来说,说明某种设计是不是"科学的"似乎是无用的。设计是研究的计划,正如在所有的研究中所表达的那样,非控制的以及控制的、主观的以及客观的。它不是一种科学的或者不科学的事情,而只有好的或者不好的设计。所要求的精确性程度、"证实"所要达到的水平、对现有知识的陈述等等,所有这些结合起来,才能决定一项研究设计中所能具有的"科学"程度的总量。

2. 假设的证明绝不是确定的。所能希望做的最好的事情是做出一系列或多或少似是而非的替换假设。在大多数情况下,会有多种解释起作用。说明自己的假设的正确性并不排斥替换假设,反之亦然。

3. 并不存在一种惟一"正确的"设计。由于偏爱自己的方法论和理论倾向,不同的研究者会做出不同的设计。假设可以根据不同的设计用不同的方法来研究。

4. 所有的研究设计都表现了一种进行社会研究时从众多实践方面考虑所要求的

妥协。我们中没有人能够不考虑有限的时间、经费和个人预算而进行研究。进一步的限制涉及资料的有效性以及研究者利用其研究对象的程度。研究设计必须是可实践的。

5. 研究设计并非是一种必须完全遵循、一点都不能走样的高度特别的计划，而是一系列使我们保持正确方向的路标。我们必须随时（尽管不是经常地）准备放弃那些无法验证的假设，同时在增加知识的基础上去发展出新的假设。此外，在现实的各种考虑面前，任何出自办公室里的研究设计都将不可避免地被改变。

资料来源：Edward A. Suchman(1954),"The principles of Research Design," in *An Introduction to Social Research*, edited by John T. Doby (New York: Stackpole),使用获得了授权。

2.14 影响研究设计内在效度与外在效度的因素

也许在过去的50年中，对于理解研究设计和计划实验来说最重要的出版物就是Donald T. Campbell 和 Julian C. Stanley 所撰写的《实验的与准实验的研究设计》(*Experimental and Quasi-Experimental Design for Research*)一书了，我们将它摘录在下面。他们关于内在效度与外在效度的概念作为重要的评价标准和警示，开启了精确地和有效地评估实验设计的大门。

内在效度是一项实验设计的这样一种质量的反映：任何结果或效果都可以归因于对自变量操纵的程度。外在效度则是实验设计的这样一种质量的反映：结果可以被概括到不同背景中的程度。很明显，二者之间存在着某种平衡。例如，我们不可能在具有非常高的内在效度的同时，具有很大的概括性。正像具有很少的控制（较低的内在效度）同时能够概括到其他样本是十分困难的一样。二者之间的平衡主要依赖于所提出的（作为假设表示的）实验问题，以及研究者所愿意冒的风险。

实验的与准实验的研究设计

Donald T. Campbell & Julian C. Stanley

列出这一清单的基础是内在效度与外在效度之间的差别。内在效度是实验的最基本要求，没有内在效度，任何实验都无法解释：实验的处置在这一特定的实验例子中确实产生出差别了吗？外在效度提出的是概括性的问题：处置变量对于测量变量的影响可以概括到什么样的总体和背景中？两种类型的标准显然都是重要的，尽管它们常常会因为一种特征的增加危害到另一种特征而发生冲突。当内在效度是必要条件，同时当外在效度的问题——如同统计推断的问题——并非完全可以回答的时候，在选择研究设计时同时增强两种类型的效度显然就是我们的理想。

有关内在效度，我们将展示八种不同的外部变量的类型。在实验设计中，如果不

对其进行控制，这些变量产生的效果将会与实验刺激的效果混在一起。它们代表了以下因素的效果：

1. 历史因素，即除了实验变量以外，在第一次测量和第二次测量之间所发生的特定的事件。

2. 成熟因素，即回答者自身随时间流逝而发生变化的过程（并不特别由于特定的事件），包括变得更年长，变得更饥饿，变得更疲劳等等。

3. 测验效应，即进行第一次测验会对第二次测验的分数产生影响。

4. 仪器因素，即在仪器校准方面的变化，或者在观察者方面的变化，或者所用的计分方式都可能在获得测量的过程中产生改变。

5. 统计回归因素，即当群体是依据他们的极端分数来选择的时候统计回归会起作用。

6. 选择因素，即对于比较组采取不同的回答者选择程序时所导致的偏误。

7. 实验的死亡率，或者比较组中回答者的缺失。

8. 选择与成熟的交互作用等等，在多组准实验设计中这种交互作用肯定会混淆于实验变量的效果之中，也可能会被错误地当作实验变量的效果。

危害外部有效性或者代表性的因素有：

9. 测验的反应性或者交互影响，即前测可能会增加或者减少回答者对实验变量的敏感性或反应性，因而使得从前测总体中所得到的结果，不能代表实验变量对抽取实验回答者的那个未经前测的总体所具有的效果。

10. 选择的偏差与实验变量的交互影响。

11. 实验安排的反应性影响，那些在非实验背景下被卷入实验的人所受到的实验变量影响的概括将会被排除。

这样一个列表的价值是，它在研究者完成一个设计之前提出了某种警告。要增加所需要的准确性程度，就不能忽视这些因素。放进研究设计的东西应该预示着当资料被收集和分析以后会有什么样的结果产出。

资料来源：Donald T. Campbell & Julian C. Stanley(1966), *Experimental and Quasi-Experimental Designs for Research* (Chicago: Rand McNally), pp. 5-6.

关于内部效度与外部效度的推荐读物

Barr, Judith T. (Ed.). (1996). *Clinical effectiveness in allied health practices: Critical literature review.* Collingdale, PA: Diane.

Braver, Sanford L., & Smith, Melanie C. (1996). Maximizing both external and internal validity in longitudinal true experiments with voluntary treatments: The "combined modified" design. *Evaluation and Program Planning*, 19, 287-300.

Durlak, Joseph A. (1998). Why program implementation is important. *Journal of Prevention and Intervention in the Community*, 17, 5-18.

Garaway, G. B. (1997). Evaluation, validity, and values. *Evaluation and Program Planning*, 20, 1-5.

Kirk, Jerome, & Miller, Marc L. (1999). *Reliability and validity in qualitative research*. Thousand Oaks, CA: Sage.

2.15 抽　样

样本是从总体中选择的一组对象。抽样的目标是抽选出一个具有最小抽样误差(即样本与总体特征之间差别)的样本。这样,样本对我们所感兴趣的总体具有最好的代表性。同时,作为科学,特别是推断统计(见第6部分)特点的推广性也最大。

抽样的问题涉及下列一些方面的影响:①总体的界定;②样本的规模;③样本的代表性。关于总体的界定,重要的问题是决定研究者希望将他的研究发现推广到哪个群体中。关于样本的规模,则必须考虑到个案持续稳定消失的情况(称作死亡率),同时应该尽可能清楚地预见到这一点。虚拟表可以帮助我们进行这种计划。此外,样本规模必须足够大,以便在统计分析中具有充分的解释力,但样本又不能大到浪费资源的地步。第三个同时也可能是最复杂的抽样问题则会产生在与获得有代表性样本的方法有关的方面,对于任何样本来说,基本的要求是尽可能对它所取自的总体具有代表性。

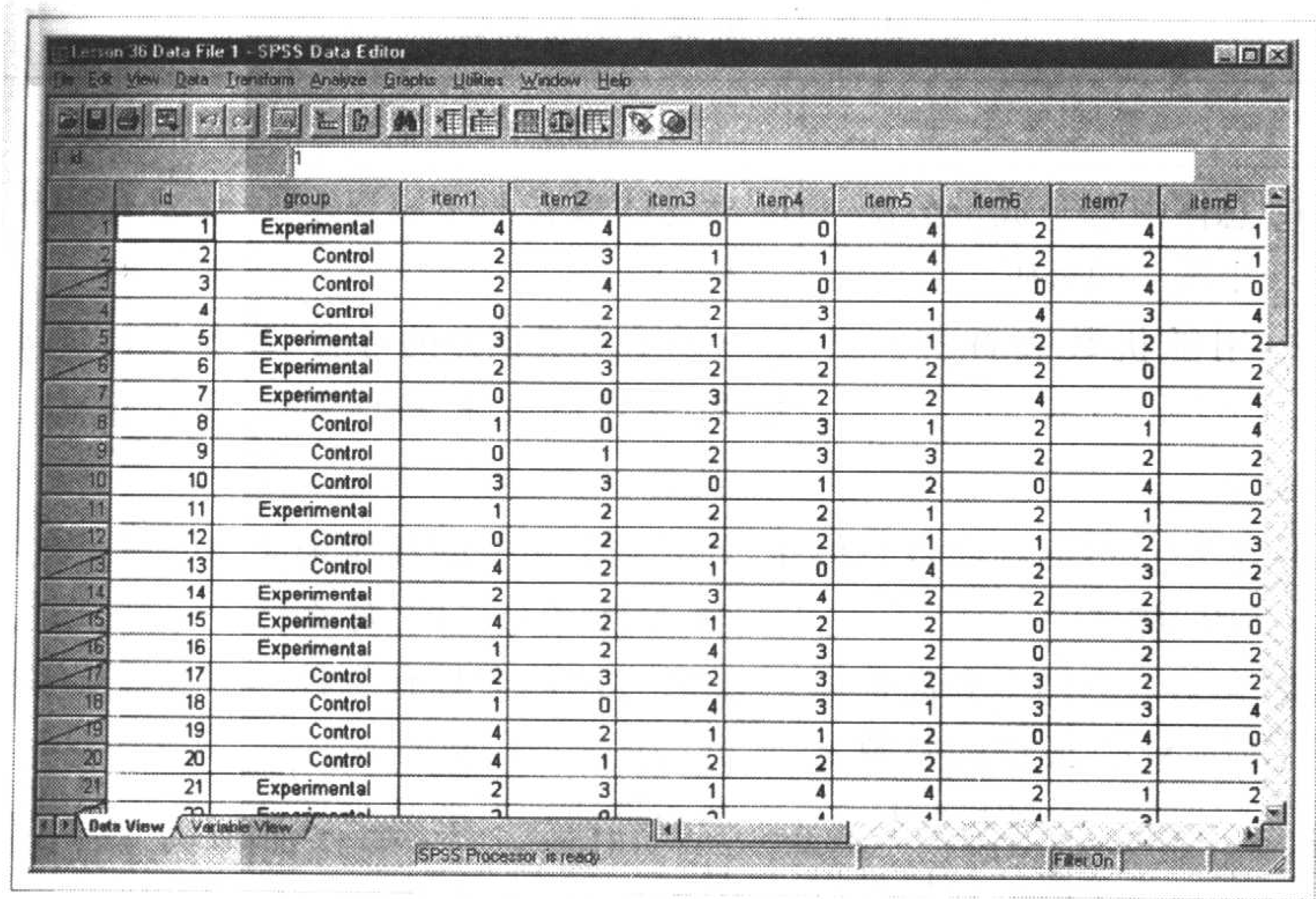
常用的三种抽样方法是:随机抽样(random sampling)、分层抽样(stratified sampling)、判断或目的抽样(judgemental or purposive sampling)。

2.15.1 随机抽样

随机抽样要求总体中的每一个成员都有同等的和独立的机会被包括在样本中。最严格的随机抽样方法要使用随机数表,即总体中的每一个成员都被给予一个数字,随机数表选择出来的数字所代表的总体成员就被包括进样本中,选择过程如此持续进行,直到抽满预先决定的样本规模为止。另一种方法是将总体中成员的名字或数字写在卡片或小纸板上,将它们搅拌均匀,然后从中抽取。

随着像SPSS这样的软件越来越普遍,选择一个样本就变得十分容易了。如图2.1所示,在数字栏中划了斜线的就是被选择进样本的个案。

系统抽样是一种方便的替换方法,它不仅与随机抽样完全相等,而且与实践目的常常更为紧密。系统抽样从总体中某个随机成员开始,每隔n个成员抽取一个。其做法是先用总体除以所要求的样本规模求出n,接着选择一个开始点,然后每隔n个选取一个。例如,如果你希望从一个100人的总体中抽取20个人的样本,你先用100除以20得到5,然后从名单上某个随机抽取的名字开始,每隔5个名字抽取一个。



	id	group	item1	item2	item3	item4	item5	item6	item7	item8
1	1	Experimental	4	4	0	0	4	2	4	1
2	2	Control	2	3	1	1	4	2	2	1
3	3	Control	2	4	2	0	4	0	4	0
4	4	Control	0	2	2	3	1	4	3	4
5	5	Experimental	3	2	1	1	1	2	2	2
6	6	Experimental	2	3	2	2	2	2	0	2
7	7	Experimental	0	0	3	2	2	4	0	4
8	8	Control	1	0	2	3	1	2	1	4
9	9	Control	0	1	2	3	3	2	2	2
10	10	Control	3	3	0	1	2	0	4	0
11	11	Experimental	1	2	2	2	1	2	1	2
12	12	Control	0	2	2	2	1	1	2	3
13	13	Control	4	2	1	0	4	2	3	2
14	14	Experimental	2	2	3	4	2	2	2	0
15	15	Experimental	4	2	1	2	2	0	3	0
16	16	Experimental	1	2	4	3	2	0	2	2
17	17	Control	2	3	2	3	2	3	2	2
18	18	Control	1	0	4	3	1	3	3	4
19	19	Control	4	2	1	1	2	0	4	0
20	20	Control	4	1	2	2	2	2	2	1
21	21	Experimental	2	3	1	4	4	2	1	2

图 2.1 展示随机抽样的计算机屏幕

2.15.2 分层抽样

前述方法假定总体的构成未知,因此通过严格的随机选择或者有规律的间隔选择可以得到最接近代表性的样本。某些情况下,在抽取样本之前,就已知总体的构成具有某些重要的特征(准确的或近似的)。极为重要的标准是,当一个总体的特征(比如性别或社会阶层)与因变量相关时,那么,样本就应根据那种特征进行分层。

例如,我们可能了解总体中男人和妇女的确切比例,而性别差异又与我们希望检验的变量相联系。在这种情况下,研究者通过选择在总体重要特征上成比例的子样本,可以增加选择有代表性样本的机会。因而他们能够选择一个在数学上绝对代表某些重要特征的样本。表 2.3 中展示了多种形式的分层随机抽样技术。

2.15.3 判断或目的抽样

当实践的考虑排除概率抽样的运用时,研究者可以通过其他(非概率的)方式来寻求有代表性的样本。他们可以寻找一个在整体上对总体具有典型性的子群体。观察限制在这一子群体中,而从所得资料中得出的结论则推广到整个总体中。一个例子是,将一个特定的州或县中选民的选择作为选举结果的一种晴雨表,将它们过去的选举结果作为其对全州或全国的代表性的依据。

这种样本不能计算其抽样误差与偏差。因此,判断抽样应该限制在以下情形中:①当可能的误差不严重时;②当概率抽样在实践上不可能时。从判断抽样所得到的资料最多只能建议或指示结论,一般不能用来作为统计检验程序的基础。

2.15.4 抽样技术综述

上述讨论的三种抽样形式并没有穷尽抽样程序的范围。表 2.3 列出了诸如多阶段随机抽样、整群抽样、分层整群抽样以及重复抽样这样一些类型。

表 2.3　抽样技术综述表

抽样类型	简要描述	优　点	缺　点
A. 简单随机	1. 总体中每个成员分配一个惟一的号码 2. 用随机数表选择样本成员	1. 对总体事先的了解要求最少 2. 不存在分类误差的可能 3. 易于分析数据和计算误差	1. 无法利用研究者对总体的了解 2. 比同样规模的分层抽样的误差更大
B. 系统抽样	1. 利用自然的顺序或总体中的顺序 2. 在第 1 个样本和离抽样比(N/n)最近的整数之间选取一个随机的起点 3. 在每一个离抽样比最近的整数区间中选取一个数字	1. 如果对于相关的性质来说,总体是有序的,那么这种方法就会具有一种分层的效果,同时比简单随机抽样更能减少变异性 2. 抽样简单	1. 如果抽样区间与总体的一种周期性的排序相关联,就会引起变异性的增加 2. 在存在分层影响的地方,误差估计可能更高
C. 多阶段随机	抽样至少分成两个阶段,在每一个阶段中都采用随机抽样的形式	1. 仅仅只要求样本中所选择的那些抽样单位中成员的抽样名单、识别身份以及数字化 2. 如果抽样单位是按地理位置来确定的,就能减少实地调查的费用(比如交通费)	1. 对于同样规模的样本来说,误差可能比随机抽样或系统抽样更大 2. 当所选择的抽样单位减少时,抽样误差增加
1. 概率与规模大小成比例(PPS)	按照概率与每个单位的规模大小成比例的方式选取抽样单位	1. 减少变异性	1. 如果在选择前缺乏对每一个抽样单位规模的了解,就会增加变异性
D. 分层抽样 1. 按比例	与随机抽样中最后阶段不同的是,从每一个抽样单位中按照这个抽样单位的规模成比例地进行选择	1. 能够保证形成单位分类基础之特征的代表性。因此,所产生的变易性比简单随机抽样或多阶段随机抽样更少 2. 通过分类过程可以减少总体成员未能被包括进来的机会 3. 能够估计每一层的特征因而可以进行比较	1. 需要有关总体各层比重的精确信息,否则就会增加误差 2. 如果没有适用的分层名单,要准备它们比较费事;存在着错误分类的可能性,因此会增加变异性
2. 最佳分配	除了按层内的变异性以及各层的规模比例抽样外,其他与按比例分层抽样相同	1. 对于同样的样本规模比按比例分层抽样的变异性更小	1. 需要了解层内相关特征的变异性

续表

抽样类型	简要描述	优　点	缺　点
E. 整群抽样	通过某些随机抽样的形式选择抽样单位,最终的单位是群体;随机地选择这些群体并将其中的每一个成员计算在内	1. 如果群体是按地理位置分布的,那么就能减少实地调查的费用 2. 仅仅只要求所选群体的成员名单 3. 可以估计群体的以及总体的特征 4. 由于所选择的是群体,而不是个体,因此可以用于后继的抽样,可以进行个体替换	1. 在可比较的规模上比其他概率抽样的误差更大 2. 要求能够惟一地将每一个成员分派到一个群体中;如果不能做到这一点,将会导致个体的重复或缺失
F. 分层整群	从每一个抽样单位中随机地选取群体	1. 减少普通整群抽样的变异性	1. 在整群抽样的缺点上又增加了分层抽样的缺点 2. 由于群体的性质是变化的,分层的优点可能会减少,导致样本对大规模研究不适用
G. 重复抽样:多重的或者相继的	抽取两个或多个上述任何形式的样本,用前面的样本的结果来设计后面的样本或者决定它们是否必要	1. 提供总体特征的估计以便于后续抽样计划更有效,从而减少最终估计的误差 2. 最终将减少所需要观察的数量	1. 复杂的实地工作管理 2. 比非重复抽样要求更为复杂的计算和分析 3. 只有当一个非常小的样本就有近似的代表性以及当观察的数量在研究的任何阶段都可以方便地增加的情况下,后续抽样才能使用
H. 判断抽样	以适当的信息为基础,通过判断,从总体中选择一个子群体作为整个总体的代表;将这一群体的全部或者其中的一部分作为样本	1. 可以减少抽样与实地研究准备工作的费用,因为可以选择那些紧密相连的最终单位	1. 不能测量和控制变异性和估计偏差 2. 很强的有关总体的假定或对总体和所选择的子群体有相当的了解
I. 配额抽样	根据相关的性质将总体分类;决定每一类中所需要的样本比例;为每一个观察者选定配额	1. 与上面相同 2. 可以引进某些分层的效果	1. 对研究对象的分类以及在每一层内非随机地选取对象会引起观察者的偏差

资料来源:Russell Ackoff(1953):*The Design of social Research* (chicago: University of Chicago Press), p. 124.

选 例

Russell Ackoff

从实践的和纯科学的目的来说,使用一个误差可测的选择程序是必要的。这种程序应该具有刻画与偏差和变异性相关的特征的能力。满足这些条件的基本程序是简单随机抽样,这是一种使每一个成员都具有同等的被选中的机会的方法。简单随机抽样依靠随机数表的帮助进行,而系统抽样是其变种,它通过在预设的区间中从一个随机的起点开始选择的程序来进行。

通过将总体分成小的群体,我们可以按如下方式在各个阶段中选择一个样本。如果在每个阶段都选择一个随机样本,我们就得到一个多阶段的随机样本。如果在一个阶段中,全部抽样单位都被抽取,而不是在最后的阶段被抽取,我们就得到一个分层的样本。如果是在最后的阶段中全部成员被抽取,我们就得到一个整群的样本。选择任何子群体的概率可以根据子群体规模的某些功能相对应地得到,而且,从任何子群体中所选取的单位数目也可以根据某些这样的功能相对应地得到。按比例抽样趋向于减少抽样的误差。分层抽样和整群抽样可以结合起来产生有效的样本。特别是当分层抽样和/或整群抽样基于地理特征时(例如按区域抽样)。区域抽样减少了准备抽样名单的复杂性,同时允许将对象整合成群以便于成群抽取。

在双重抽样中,第一个样本可以用来提供信息,这种信息进而又可以用来设计一个有效的第二样本。通常,在通向研究结论的过程中,这种抽样也能够用来减少所需要的观察数量。当双重抽样被推广时,它会产生后续的抽样,即一种一次抽取一个或一组个案,同时运用所得到的资料来决定是否继续进行抽样的方法。

选择一种抽样程序的根本基础应该是,使获得样本的费用最小化,同时使由运用方法所产生误差的预期费用最小化。在进行这样的评估时,应该寻求专家的帮助。

上述抽样图表(表 2.3)是以非常简要的方式对所讨论的各种抽样程序的基本描述,对各自优点和缺点进行了总结。

资料来源:Russell Ackoff(1953):*The Design of social Research* (Chicago: Chicago University Press), pp. 123-126

关于抽样的推荐读物

Abbott, Andrew, & Barman, Emily. (1997). Sequence comparison via alignment and Gibbs sampling: A formal analysis of the emergence of the modern sociological article. *Sociological Methodology*, 27, 47-87.

Handwerker, W. Penn, & Wozniak, Danielle F. (1997). Sampling strategies for the collection of cultural data: An extension of Boas's answer to Galton's problem. *Current Anthropology*, 38, 869-875.

Maisel, Richard, & Perseli, Caroline H. (1995). *How sampling works.* Boston: Pine Forge Press.

Rosnow, Ralph L., & Rosenthal, Robert. (2001). *Beginning behavioral research: A conceptual primer* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Valliant, Richard L., Dorfman, Alan H., & Royall, Richard M. (2000). *Finite population sampling and inference: A prediction approach.* New York: John Wiley and Sons.

关于基础研究与研究设计的推荐书目 2.16

Ackoff, Russell L. (1953). *The design of social research.* Chicago: University of Chicago Press.

Ader, Herman J., & Mellenbergh, Gideon J. (1999). *Research methodology on the social, behavioral, and life sciences: Designs, models, and methods.* Thousand Oaks, CA: Sage.

Alwin, Duane F. (Ed.). (1978). *Survey design and analysis.* Beverly Hills, CA: Sage.

Armer, Michael, & Grimshaw, Allen (Eds.). (1973). *Comparative social research: Methodological problems and strategies.* New York: Wiley.

Blalock, Hubert M., Jr. (1984). *Basic dilemmas in the social sciences.* Beverly Hills, CA: Sage.

Blessing, Lusienne T. M., & Chakrabarti, Amaresh. (1999). *A design research methodology.* Godalming, Surrey, UK: Springer-Verlag.

Bordens, Kenneth S., & Abbott, Bruce B. (1998). *Research design and methods: A process approach.* Mountain View, CA: Mayfield.

Brinberg, David, & McGrath, Joseph E. (1985). *Validity and the research process.* Beverly Hills, CA: Sage.

Fielding, Nigel G. (1988). *Actions and structure: Research methods and social theory.* Newbury Park, CA: Sage.

Grunow, Dieter. (1995). The research design in organization studies: Problems and prospects. *Organization Science*, 6, 93-103.

Hakim, Catherine. (1987). *Research design.* Winchester, MA: Allen & Unwin.

Jensen, Paul A. (1999). *Models and methods in operations research.* Upper Saddle River, NJ: Prentice Hall.

Kenny, David A. (1996). The design and analysis of social-interaction research. *Annual Review of Psychology*, 47, 59-86.

Kuehl, Robert O. (1999). *Statistical principles of research design and analysis.* Pacific Grove, CA: Brooks/Cole.

Lieberson, Stanley. (1985). *Making it count: The improvement of social research and theory.* Berkeley: University of California Press.

Strauss, Anselm, & Corbin, Juliet. (1990). *Basics of qualitative research: Grounded theory procedures and techniques.* Newbury Park, CA: Sage.

3

APPLIED AND EVALUATION RESEARCH

应用与评价研究

第 3 部分描述了应用与评价研究者如何面对操作型或政策决策型的问题。不同于基础研究,这种研究问题并不常常被研究者主动选择,但却常常由于管理者或立法者的即时需要而被摆在研究者的面前。

应用型研究者忙于为面临的问题提出解决办法,而评价型研究者评测一个项目或一个过程的最终结果。这两种类型的研究具有非常大的挑战。在本章中,我们把关注点放在应用和评价研究者必须思考和解决的技术性和社会性问题上。费用和效力是评价研究者工作的常用标准,正如你在本章后面部分将要看到的,它们常被用来测评一种政策的有效性。

一条与政策决策相关的应用社会科学基本原理 3.1

由于在拥有严谨的学术市场的大学之外，就业机会开始增多，应用社会科学重新获得了广泛的关注。而且，在过去的20年里，人们发现社会科学正起着越来越重要的作用，广泛地支持着各种类型的社会项目。

作为国家科学基金会(National Science Foundation)的社会领域和经济领域的先驱，Otto Larsen对社会科学家在政策决策方面的作用这样解释道：

> 由在任的官员，而不是科学家决定公众政策。但通过各种各样的机构，来自社会和行为研究领域的研究，可以并且切实地在影响决策制订的过程。如，国家科学院(National Academy of Sciences)的国家研究委员会在国家政策方面经常被咨询相关意见。他们着重运用社会和行为科学研究成果，通过它们来评价项目以及处理社会上比较关注的事情，如，能源、税收、生物医学的组织构造学、环境监控、酒精滥用、个人隐私保护、人口老龄化、噪音消除、儿童发展、人口生产和死亡率的改变等。在处理诸如暴力、淫秽、人口或犯罪等许多重要事务方面，情况也是如此。政府之外的组织也用社会科学的数据去影响和建议政策的决策过程。比如，在斯坦福大学胡佛学院的大力支持和赞助下，以经济学家和政治学家为主的一批杰出的学者提供了对一本名为《20世纪80年代的美国》(*The United States in the* 1980*s*)的书中许多主要的国内国外观点的相关评论和分析，这本书由Peter Duiganan 和Alvin Rabushka编写。(Larsen，1981，p.8)

Peter Rossi认为，通过在应用社会科学协议上的竞标，社会科学部门有了很好的机会去实现作为社会科学专家所能从事的服务："作为一个有组织的学科，我们必须与应用科学世界建立联系，使协约机构和研究行业知道社会学可以提供一些建议和帮助，并让我们自己和学生也知道，应用社会科学是一个很有趣并让人兴奋的职业。"(Rossi，1980，p.20)

美国应用社会科学职业机会委员会的主席Howard E. Freeman针对目前应用社会科学的现状在自己的著作中提出这样的观点：

> 在学科领域中，应用社会科学有自己很深的根基。从20世纪30年代开始，很多优秀的社会科学家提出要把社会科学的发现成果应用起来，很多会议谈到应用工作的重要性，还有许多书都证明了社会科学研究的有用性。而且，随着校园以外的工作和就业机会、为研究提供支持的可能性、毕业生培训机会的不断增长，应用社会科学在学科内的成长也随之被证明。"1980年社

> 会学系研究生指南”列出了一百多个系开设了应用社会学方面的课程和特殊项目；一些大型的营利和非营利的研究组织所雇佣的社会学者比许多社会学系雇佣的社会学者还多；而且联邦对基础研究的支持只占应用性研究基金的一小部分。（Freeman，1980，p. 1；也可参见：*Study Project of Social Research and Development*，1978）

相应地，所有应用社会科学的培训项目必须明确，要影响政策决策者的行为，有很多因素必须同时考虑。Leonard Saxe 写道：

> 政策决策者能够成功运用社会科学的关键依赖于很多因素。这些因素包括问题是如何定义的，研究的可信度，研究者所引发的争论的性质，研究证据的质量和可获得性，研究者交流他们思想和研究成果的能力，以及政策决策者恰合时宜地关心某个论题。（Saxe，1987，p. 226）

Stan Capela 是 2000—2001 年应用社会科学家协会的会长。他为应用社会科学这个领域在新千年的头十年发展方向提供了如下视野。

应用社会学：从这里走向哪里

Stan Capela

这是有关应用社会科学的一系列重要专栏的开始部分。每个月我将花一些时间来探讨应用社会科学的现状，同时看看它应在事情计划中的哪个位置。

在近来的联合会议上，有许多关于 21 世纪应用社会科学将何去何从的问题讨论。为了回答这个问题，首先我必须判别清楚这个问题和其他三个不同群体的关系。第一个群体是活动于学院外的实践型社会学家；第二个是包括在学院内的那部分社会学家，他们必须教授学生和给学生们提供面对开放的市场所必需的技术；第三部分是学生，他们正在思考社会科学与他们毕业后希望从事的工作是否具有很强的相关性，是否值得他们去学习和研究。

假如我面对实践型社会学家来做演说，我就假设他或她可能会想知道应用社会科学家将来的工作立足点在哪里。是从事研究？是去做一个评价者？还是将在政府里工作？假如我对学院内的社会学家讲演，我可能必须提供一些信息，在将来什么样的技术是需要的，以及在教室里这些信息可以如何用最好的方式去解释。至于对学生们，他们想知道他们的工作在哪里。

在“同心人文服务机构”做管理信息系统（MIS）主管期间，我发现，应用社会科学家将来的发展方向是成为便利者。特别是，当他们工作于组织化程序很高的系统中时，CEO（执行总裁）往往会求助某个人，希望他能澄清困境并提出解决方法。很多例子中都表明，CEO 希望你能集合一个群体的成员，这些人有自己的关注热点，你必须能使他们成为一个群体来面对问题的解决，并最终达成一致意见。很多实例表明，你

必须处理好这样一部分人:这些人有各自不同的行事主则,他们认为任何一种解决办法都会对他们产生消极的影响,并且他们会把这种不佳的情况归咎于他人。

应用社会科学家在这方面能很好地工作,因为他们拥有群体工作的技巧。第二,他们明白语言在意见上达到一致是非常重要的。更重要的是,他们收集信息和分析情况的理论基础给他们自己提供了一个温和的方式,这种方式建构了达成一致意见的基础。在这种最终的分析中,假如应用社会科学家能够使他们自己的理论基础成为一种根基,假如他们以一种容易理解的方式提出信息。那么,在各种组织性的机构中他们可以发挥长远的重要作用。

你是如何想的呢?你如何发展这种观点?提升你的观点吧,这样我们可以有希望构建出这种观念,最终,它将满足我们这个领域不同需求者的需要。

资料来源:Stan Capela(2000),*Applied Sociology: Where Do We go From Here*, Newsletter, Society of Applied Sociologists,使用获得了授权。

参考文献

Freeman, Howard E. (1980). *Footnotes* (American Sociological Association), p. 1.

Larsen, Otto. (1981, March). Need for continuing support for social sciences. *Footnotes* (American Sociological Association), p. 8.

Rossi, Peter. (1980, August). *Footnotes* (American Sociological Association), p. 20.

Saxe, Leonard. (1987). Policymakers' use of social science research. In William R. Shadish, Jr., & Charles S. Reichardt, eds., *Evaluation Studies Review Annual* (Vol. 12). Newbury Park, CA: Sage.

Study Project of Social Research and Development. (1978). *Study project report: Vol. 1. The federal investment in knowledge of social problems.* Washington, DC: National Academy of Sciences.

应用研究设计 3.2

与基础研究相比,应用社会学研究有很多重要的差异,它有许多新的规则和要求:

- 应用研究必须忙于研究由某位顾客提出的具有紧迫性的问题(公众的或私人的)。这个顾客可能是研究者在商业、政府或服务组织方面的老板,或者这个顾客可能是外部组织正在寻求的专业服务,或咨询管理者,或研究代理等。这种顾客经常是一个政府或私人的基金代理。他们需求同样的东西:一种可以用来回答政策决策者所面临问题的有效知识。
- 研究者必须是一个很好的转换者,能在他从事的学术原则与现实世界行为中很好转换。转换中存在的主要要求来自于现实世界行为,包括合时性、行动、每日用语和概念的使用、过去现在涉及的特殊兴趣、冲突、在资源上的争斗等。(Coleman,1972)

- 研究者必须判断，找出行为结果和理论之间的关系是否重要。学术研究者经常在相当大的压力下出版他们具有科学观点的成果。而基于指导理论的验证假设模式必需接着出现。这种要求将会影响基础研究标准的设计，但这种要求却对现实性政策指导有一种很强的积极推动作用。
- 这种研究是由顾客支付费用的，它的研究效果是由它解决或减轻管理者（或政策）问题的程度来评价的。

已故的哈佛大学 Samuel Stouffer 教授指导过对美国二战军人的大规模研究，他曾说，“在为美国军队进行可操作性的研究时，假如我有机会去做 10% 的基础研究，我将是非常幸福的。”他在这两种类型研究上的成功通过美国军队的 George Marshall 将军对他的称赞可以证明。同样他也因为写了四大卷的《美国士兵》（*The American Soldier*, Stouffer, 1947, 1949）被美国的社会学家和心理学家称赞。作为联系基础和应用研究的桥梁，《美国士兵》是一种典范。

无论是基础的还是应用的研究者都需要资金，但相对于基础研究来说，更多的资金拨给了应用研究，因为联邦政府在它所支持这些研究上需要有好的统治政绩。换句话说，应用研究的结果经常回答了直接的和明显的问题。此外，这种能够通过提供研究设计把基础研究和应用研究结合起来转化成资本的能力对政策决策者和重要人物而言是非常渴望和具有吸引力的。

参考文献

Coleman, James S. (1972). *Policy research in the social sciences.* Morristown, NJ: General Learning Press.

Stouffer, Samuel A. (1947, 1949). *Studies in social psychology: The American soldier* (4 vols.). Princeton, NJ: Princeton University Press.

3.3 为一个问题设计应用研究计划

应用研究计划的设计必须能够解释复杂的社会系统中的行为。研究者一开始就要试着去将研究系统中的各个部分，以及它的范围、它与其他系统的相切点、它的反馈路线、它可能联系的其他子系统等等概念化。关键的问题主要有以下两个：

- 我们如何在系统中获得明确的结果？
- 涉及哪些关键性的原因要素？

分析是从对情境因素的定位开始的。它的过程是以清晰化存在的心理和行为反应为内容进行的。人与人之间有很多因素，比如感觉、规范、价值、评价和目标等，这些都必须作为参与的相互作用因素放进程序图中加以考虑。接着，介入理论必须设计成可操作利用多种情境因素，这是政策和结构变化的结果。在这种介入后，评价就可以开始了。

Chris Argyris(1985;Argyris,Putnam,&Smith,1985)已经指出,开始做这件事情的人需要考虑何种类型的知识在行动时是个体所需要和使用的。个体对他们世界的感知是通过把所拥有的信息数据组织成各种类型的模型,存在大脑里,在任何需要的时候提取它们。社会科学家可以帮助确保个体生成的知识,通过行为程序图来促使这些行为模式变得非常有用。

Argyris 制作了程序图,在图中设置了很多在给定的问题情境中会出现的因素,同时当情境中出现的参与者们有合作和冲突时对他们进行追踪。这些步骤的结果包括了参与者在取得成功和他们试图解决失败时面临的压力,以及他们随之而来的行为方式。

参考文献

Argyris,Chris.(1985). Making knowledge more relevant to practice: Maps for action. In Edward E. Lawler III & Associates, *Doing research that is useful for theory and practice* (pp. 79-106). San Francisco: Jossey-Bass.

Argyris,Chris,Putnam,Robert,& Smith,Diana McLain.(1985). *Action science: Concepts, methods, and skills for research and intervention.* San Francisco: Jossey-Bass.

学院的研究者在应用研究中的心情 3.4

大学教师在应用研究方面存在着一种矛盾的感觉。J. Richard Hackman(1985)这样解释它:

> 首先,尽管我个人有理由充满信心地认为,新概念上的和方法论上的方法在组织行为中是必需的,但在偶尔使用这些途径的实验中,我会因为我与追寻科学价值的传统方式发生偏离感到自己是异教徒以及有罪恶感。第二,当一个人试着去做某件他并不知道该如何去做的事情,同时也没有现成的模型存在时,失败是很可能存在的,并且失败的比率远大于成功。这种矛盾开始带来了焦虑。接着,这种焦虑阻碍了工作很好地进行,使得要做好任何事情成为一种困难,更不要说去做某些新的或是有趣的事了。但是,矛盾和焦虑常常被称为是创新的先驱,因此,人们总是希望,假如坚持足够长的时间,某些有价值的事情也将出现。
>
> 当然,多种疑问和观点是会出现的,它们存在于情感方面和理性方面,当我试着继续发展应用性的有关个人的和群体执行效率的理论,这些问题便一直困扰着我。当然,我发现这样做对我的想像力和勇气都是一种潜在的挑战。(pp. 148-149)

参考文献

Hackman, J, Richard. (1985). In Edward E. Lawler III & Associates, *Doing research that is useful in theory and practice*(pp. 148-149). San Francisco: Jossey-Bass.

3.5 应用研究在政策分析中的成功运用

作为社会与行为研究的工具，政策分析的使用可能是应用研究中最成功的例子之一。在这里，基础数据被使用并应用到现实世界的政策性问题里，这些数据是由一整套的测评标准来评价的。下面的内容是这样一个评价（集中在对日间护理选择的讨论上）的摘录，包括这些评价性的指标和他们是如何使用的。

政策分析

Ron Haskins

在使用工具去帮助判断某种公共政策或其他政策的相对优点方面，近几年有一种非常流行的做法，这就是运用政策分析。尽管政策分析并不是一个新的领域（MacRae, 1976; Nagel, 1977），但作为政策分析工具应用于社会问题的做法，即使不是新的，至少也是被心理学家、教育学家和其他社会科学家逐渐关注起来的。在本卷中作为论文来论证的是，在由研究者和其他专业人员（这些人都专注于对孩子及其家庭研究）组成的大的、并在某种程度上存在分歧的群体中也都开始把关注点直接放在自己研究领域的公开政策含义上。本章介绍的目的在于试图勾勒出一个详细的方法，这种方法可以用来检查由社会政策所试图解决的问题。

什么是政策分析？

MacRae 和 Wilde 把政策分析定义为"在众多可选的政策中使用原因和证据去选择最好的政策"（p. 4）。像其他定义一样，这种定义（Rein, 1976; Titmuss, 1974）强调了在政策选择中原因和证据的使用。在这些考虑中，政策分析和社会科学存在很多共同点。原因的使用意味着决策是根据一系列清晰的标准进行的，这些标准本身被随手可得的事例证明是恰当的。于是，稀奇古怪的想法、个人的观念以及政治信仰都必须在决策中被最小化，且受制于逻辑的和推论的规则。

要克服个人想法和偏见的一个重要因素是在强调逻辑时，还要在众多的可选政策中选择好证据应用。尽管社会科学研究组成了一种重要类型的证据，但它决不是分析家惟一可用的类型。社会问题的历史常常可以提供证据，这些证据应该被考虑到，正如人口统计信息所能达到的那样。分析家若是只中意某种理论学科，就会受到限制，什么都无法提供。事实上，对于诸如健康、日常护理、收入维持、教育等大众的社会政策问题，其中任何一个都不可能由一门学科就可以提供所有相关证据。

举个例子,比如儿童护理,在美国任何一种关于这个问题和潜在解决方法的适中的、完全的分析都要考虑儿童护理的历史,提供一个关于美国儿童护理市场现状的经济描述,检查社会学上和心理学上的各种儿童护理对儿童和家庭的影响。这样的一种分析必然是跨学科的。"跨学科的",是另一个现在比较流行的词。它指的是一种从不止一个学科那里吸取信息和技巧的行为。这个术语融合了多种语言,但是,当发展学家开始涉及收益—成本分析、贸易、外部事物时,经济学家却开始考虑早期的历史体验、侵犯性的发展,以及基因—环境相互影响的复杂性,跨学科间的行为呈现出一种具体的意义,专业人员很快就认识到证据和技巧的深入性和重要性,而他们的技巧几乎没有受过正规训练。但是假如证据拒绝认可学科间人为的界限,分析家们只好扩展自己的观点并克服本学科的偏见。

一个政策分析模型

对政策分析详细定义了之后,现在我很乐意略述一种可以用来分析的特殊方法,这种方法可以保证成功合理地抨击社会问题。尽管给出的框架很简要,但通过它,我期望能提供一套具体的建议,这些建议放在一起就可以组成一种简单但却涵盖广泛的方法。这种方法可以用来分析社会问题,选择解决方案,并同那些可以颁布和完善这种方案的人交流这种方案。

这个模型由五个相互独立的步骤组成:分析问题境况,把分析时用的标准具体化,产生出可以备选的策略方案,把用于选择备选政策的信息合成起来,以及检测这种备选策略被选择的可行性。在对这个分析模型的每一步骤进行详细阐述时,我都将运用关于日间托儿政策的相关例子。尽管例子可以从各种政策问题中去提取,但日间托儿是最好的选择,不仅因为它是适时的,而且,在模型的每一步它都可以提供非常详细和丰富的说明。

问题境况的分析

很多问题的出现都源于社会人口特征的变化,传统行为模式的打破,进化出一种新的价值观和生活态度,或者技术革新的出现。这些力量出现所带来的结果就是:某个群体把这个看成是前进的机会,或者认为目前的状态受到威胁,或者感觉这是一种无法实现的需求。因此,在此处分析中,做个历史的回顾是非常必要。事实上,有人会认为不用去检验问题的起因,或者即使能获得处理问题的初始步骤也不用了解它,但这样使得要理解问题的现状成为不可能。通过历史可以获悉现在,并且,这样做的同时,历史还可以提供重要的洞察力来察觉政策问题的种类和它潜在的解决方法。问题境况的分析还必须包括对那些希望实施特殊政策的群体以及被政策影响了的群体的回顾。这种回顾将能显示出,经常是那些最有关的群体看到一种可能来临的价值或者一种能带来重要价值的机会。

对于很多政策性问题,详细说明问题境况还应该包括对那些会影响现状的现行州政权和联邦政策的研究。对家庭和孩子发生影响的问题几乎都被州政府和联邦法律规定了。于是,早先的和现在的政府活动就构成了问题境况的一部分,如果分析家们要避免提出的政策和另一个先行的政策相违背的尴尬局面,就必须把它们考虑进来。

进一步说,检测现行政策可以获得在修改新政策时必需的非常重要的洞察力,这在后面将会讨论到。

最后,彻底地分析问题境况可以对问题有一个清晰和全面的陈述,这恰是分析家目的所在。相比于分析家们在回顾这个问题境况之初时的见解,这种陈述是非常不同的。事实上,对于分析问题境况而言,一个特别重要的理由是,一种全面的分析经常能改变分析家对问题、问题的起因、问题潜在的解决方法的认识。这种改变造成了对问题的重新陈述,而这种陈述处在同历史事件更一致以及使得那些对政策感兴趣并产生影响的群体对问题有更全面了解的状态中。

产生备选策略方案

现在分析家处于产生备选方案的这一步。这个阶段的分析目的是确认所有可能用于去解决问题的方案框,并同时把那些看起来并不可行的方案去掉。

要选择备选的策略至少有四种思路来源可以操作。其中最重要的一种是历史。正如在这之前已说过的,历史背景将可以提供许多备选的策略方案。它们中的一部分可能的确已经被某种水平的政府尝试过,有些可能被某些个人或组织建议但永远不能实施,还有一些可能会被熟悉政策问题和政策背景的分析家发现。第二种思路是看看其他国家的实践和政策。尤其是要关注社会性项目,比如儿童护理、收入支持,以及健康等问题,许多欧洲国家在这些方面长期以来具有进步和有效的政策。(Kamerman & Kahn,1978)

社会科学研究也可以对备选策略给出一些提示。在这方面特别显著的例子是收入维持实验(income maintenance experiments,New Jersey,Gary,Iowa/North Carolina,and Seattle/Denver),这个实验提供了关于保证收入的效果方面的丰富信息。从这些研究中所发表的论文涉及有工作的母亲在儿童护理中的作用、社会性服务的作用、儿童的学业成就、健康保险方面的消费等内容。这些类型的信息可以用来测评收入维持策略的效果如何。在每一个相关的政策中,这种方法被认为是一种对待各种社会问题的基本解决办法。(Keniston,1977;National Academy of Science,1976)

在产生备选策略时分析家必须检查的第四种思路是来自于特殊政策和有权势的公共人物。这些特殊政策可能被利益集团、杰出的研究或政策组织所欣赏,而公共人物要包括政治家。被个人和组织所支持的可供选择的政策会在公共演说、书面报告或宣传册子中通过特别的细节体现出来。

一旦这些备选策略被确认,分析家就会试图去减少一些备选策略。因为这些策略中有些花费太大,或不切实际,或被当权的个人或群体极力反对。无论如何,形成最后的备选策略名单属于政策选择的最重要阶段。

具体化评判标准

正如Quade(1975)和其他人所指出的,政策分析产生于系统研究、操作研究以及公共事务管理。已经应用了这些分析方法的许多问题是有关科技的,比如,建造高速公路、指挥战争以及把人送到太空等等。现在各种群体和个人在社会问题上应用这些方法,尤其是在儿童和家庭政策上,非常多的困难也随之出现。这些困难出现的根本原因在于人类社会问题的解决牵涉了各种价值观的合成。简单地说,各种群体进入政

策领域时带有各自的竞争需求、利益、主张和权利。

那么如何解决这些不能以同单位度量的情况呢？正如我将谈到的，通过使用评判标准，社会政策分析家可以回答这个问题。在这里要说的最重要的一点是：在分析儿童和家庭政策时，评价必须是发挥了强有力的，甚至是卓越的作用。因此，分析模型一个最重要的特点在于它是某种可以使评价更清晰化的方法，并且，在需要的时候，可以在具有竞争性的评价中进行选择。这就正如 MacRae 和 Wilde(1979)所提出的，假如政策分析是一个"使用原因和证据选择最好的政策"(p.4)的过程，评判标准就是政策分析的核心，因为它们是我们判断备选策略是否该用的码尺。

许多这样的评判标准都是通用的。比如，它们可以并应该在分析许多社会政策时被用到。四个有关这些通用标准的示例是：水平的和垂直的公平、功效、优先选择满意事物和负面印记。

水平的和垂直的公平(horizontal and vertical equity)。水平公平包括对相同层次群体的同等对待。举例来说，所有的儿童都有权利享受充足的健康护理和教育的权利。垂直公平是指对不同层次个体和群体的分层对待。为了满足垂直平等的标准，一个政策必须引起社会资源的重新分配，以便可以给穷人提供比富人更多一些的利益。就像大部分评价标准一样，水平的和垂直的公平标准并不被所有的政策性组织作为主要的目标来接受。自由主义者和保守派认为，在财政报酬上，社会应该有一些公平的实质性的差异存在，这样可以维持增加生产率的激励机制。

功效(efficiency)。第二个重要标准是功效，它并非完全独立于垂直公平。功效是指在使用资源时产出利益达到最大化。运用这种标准，假如一个政策大体上可以产生更多的利益，那么另一个政策被用到的可能性就较小。很明显，利益这个词在这里没有很清楚地限定。什么是利益？谁得到利益？可以用怎样的单位来测量？

经济学家想出了一个方法可以部分地来回答这些问题。通过用美元表达一个项目利益和成本，利益—成本分析可将功效最大化。除了一些复杂的因素，诸如机会成本(MacRae & Wilde,1979)，项目费用测量通常是非常直接的。因此，我们经常能够用美元测量对于一个给定的社会项目制造了多少费用，或者计划或估算多少费用。但是测量社会项目的利益通常困难得多。用什么样的单位我们可以测量通过学前教育项目产生的 10 分制 IQ 的价值？通过训练项目把 1 000 个妇女投入工作的价值？从贫困家庭选出 1 000 个孩子获得一顿公平的午餐的价值？婴儿死亡率减少 10% 的价值？

功效的标准在所有的政策决策中应该扮演着非常重要的角色。因为当公共基金被有效地使用而每个人在这种长期的过程中都有所得时，任何人就没有理由对这个标准提出异议了。然而，必须承认，许多社会政策所期望的后果是难以测量的。这个难题限制了功效标准运用的能力。

满意优先(preference satisfaction)。在一个自由受到重视，且为了尊重个人的选择要尽量避免政府干预的社会里，满意优先必须被看作一个基本的准则，因为当政府帮助个体满足了自己的偏好时，个体的状态就会变得非常好。在涉及与效率有关的事物时，大多数人同意这条标准是非常重要的。

不过，至少有两件复杂的事使得这个标准很难最大化：第一，如果个人可以购买他

们所喜欢的商品和服务时,偏好可以得到最好的满足。但是,在我们社会,许多个人和家庭都没有足够的钱去满足他们的一些基本偏好。第二,很明显的是,不是任何人都有相同的偏好,因此,一些个人和团队的喜好会有冲突。

负面印记(stigma)。这个词用在这里是表示,政策产生于那些被某种否定的方式所标记,从而使他们不同于那些不受该政策影响的人们。这种标准所隐含的价值是,如果不是某些市民具有身心不健全、体弱多病,或者对参与者的病理特征,所有的市民都有权利去从政府的项目中受益。

因此减少负面印记看起来是明智的。这不仅仅是因为它妨碍了家庭的隐私,还因为它会延长一个特定的政策起到改善作用的时间。尽管具有说服力的证据还不能获得,但有理由说当父母或孩子持续地被提醒他们处于下层以及对政府救济的依赖时,他们会怨恨这种对待,他们会认为自我尊重受到了损害,进而发展为有一种无望感(例子可参阅 Rainwater,1970)。无论如何,去冒险尝试政策所带来的后果,而这种后果本来可以避免时,是极不明智的举动。

政策细节标准

除了一些刚才谈到的普遍的标准,每个政策问题都将会有特殊的标准以适应特殊的问题。这些标准是一些典型的类型,它们和某个特定政策的特殊目标相关,它们可以与那些用来解决问题的特殊策略或备选策略相对照。

打个比方说,为了更好地说清这个类型的标准再来谈谈日间托儿政策。我们可以发现一系列有关这个政策的重要纬度可以用来比较。拿两个常见的标准类型来比较,它们每一个都有很多细节标准。这两个标准都在儿童的成长中对儿童的护理发生作用,并对家庭发生影响。先看前者,分析家可能考虑它发生在个体发展、社会发展以及健康上的作用。对于后者,分析家可能考虑其发生在亲子关系中、母亲职业、婚姻满意度和母亲成就感上的影响。当然,这些并非是测量日间托儿政策效果的惟一可能标准,但是,它们表现的是分析家对身边政策问题评判时所应该包括的内容。

除去这些普遍和特殊的标准,另外一个标准值得注意。几乎每个政策都会有事先没有计划或没有预料到的效果发生。由政策引发的这些结果或者是积极的,或者是消极的。不管何种情况,有经验的分析家都可能事先预料到,并且避免它或把这种伴随的结果转化为资本。

可行性(feasibility)。一个政策将被真正实施的可行性或可能性已经在产生备选策略阶段扮演了重要的角色。在产生了一大串让人耗费精力的政策选择时,它常常被建议给分析家,要分析家们运用它作为标准去裁减那些明显看起来不可行的选项,这样可以避免对这些选项更深一步考虑以减少麻烦。

在做了一个政策选择后,可行性又一次成为主要需要考虑的事项。事实上,如 MacRae 和 Wilde 提到的,当政策选项已经被选过后,很多工作仍然需要去做。特别且有必要对将要颁布的计划和准备执行政策的规划进行修订。

颁布(enactment)。选择了最好的政策选项后,分析家就要考虑如何使这个选择得到立法的支持。为了做好这个工作,分析家在很多情况下成为了一个倡导者。当然,倡导者的角色是否管用很大程度上取决于分析家的社会地位,因为市民或分析家

都可以经常性地轻易成为倡导者,而受雇于公司或政府的分析家可能并不容易做好这个角色。

执行(implementation)。就算一个政策颁布了,它发生在孩子和家庭身上的切实效果在政策执行之前还无法呈现。优秀的政策、充足的基金可能会由于联邦或州政府的水平以及传递服务机构的传递场所而无法发挥作用。

首先,可能最重要的一点,是通过有关行政法规和指导的书籍把立法指示转变成政策。这些指导代表了一些意见,这些意见决定大批资金如何运用,报告如何书写,服务如何传递。因此,分析家必须仔细关注这些指导是如何写成,如何准备好在需要的时候提供帮助,以及在必要时抗议指导中的明显错误。

在指导中最可能出现问题的是选来执行政策实施的代理管理机构。在大部分情况下,这个代理管理机构在编写立法的过程中已经扮演了非常重要的角色。但即使如此,代理机构可以通过给其工作人员在立法执行过程中赋予的最大工作弹性来编写和改变指导。因此,如果代理机构很敌视这个政策的话,立法的目的被改变就不会让人奇怪了。事实上,一个当权机构要想扼杀一个他们所反对的政策,只需要选择一个也反对这个政策的代理机构就可以做到。

第三个在执行过程的严重问题是仇视这个政策但又没有足够政治力量去阻止政策的个体或群体可能颠覆它的执行。可以用来达到这个目的的方法有很多种,包括刚才谈到的两种。此外,国会中的反对派可以通过阻止政策基金的合法性或权威性使得政策的执行被耽搁。

第四,很多政策在执行时失败是因为能够保证成功的方法或知识并不具备。

政策选择:信息的合成

现在各种备选政策可以通过评价它们的效果来进行比较,正如可以通过每个标准来测量它们一样。分析家运用社会科学研究、历史信息,以及最好的推测来对它们评价。

当然,量性评价是首选,但是鉴于目前的知识现状,评价时要达到精确的程度常常是不可能的。所以,必须使用一些在社会科学中非传统的方法。关于这种方法举的例子是特尔菲方法(delphi procedure. Linstion & Turoff,1975)。在这个程序中,专家被要求去评价关于某个特殊的政策在某个变量上的量化评价。这些专家单个人做的评价各不相同,构成评价群。接着,他们又被要求做第二次量化评价。明显地,这种程序试着鼓励在专家们之间意见达成一致——而这其中潜藏的一个假设是,在专家们之间只有对未知的变量评价测量达成一致的看法才是最好的测量方法。

我们可以参考决策矩阵(以下图表中列出了通过用刚才讨论的评价标准来对各种备选政策做的比较):

	水平公平	垂直公平	功效	偏好满意度	负面印记	可行性	颁布	执行	政策特殊标准
现状									
备选政策 1									
备选政策 2									

第一个备选政策总是用来维持其现状。MacRae 和 Wilde (1979)建议,他们所参考的那种备选政策好像什么也没做,实际上应该包括所有的分析,这是和其他政策进行效果比较时所必需的一条基线。当然,我们必须牢记,事实上任何一个会对孩子和家庭产生影响的政策问题,为维持其现状时所做的选择不能没有政府的干预。实际上,各级政府常常被卷入各种对孩子和家庭有影响的事件中。因此,其现状的备选方案意味着目前政府政策的继续执行。

资料来源:Ron Haskins(1980), *Care and education of young children in American policy, politics, and social science*(Norwood, NJ: Ablex),使用获得授权。

3.6 关于应用社会学、知识应用、政策制订及评价的推荐书目

3.6.1 应用社会学

Anderson, R. J., and Sharrock, W. W. (Eds.). (1984). *Applied sociological perspectives.* New York: Routledge.

Birchenall, Joan. (1998). *Sociology applied to nursing.* Philadelphia, PA: W. B. Saunders.

Brown, William R. (1984, August). *An assessment of teaching applied sociology.* Paper presented at the annual meeting of the American Sociological Association, San Antonio, TX.

Clute, William T. (1986). Student consultants: Teaching applied sociology in substantive courses. *Teaching Sociology*, 14, 196-199.

Freeman, Howard E. (Ed.). (In press). *Applied sociology.* Ann Arbor, MI: Books on Demand.

Freudenburg, William R., & Keating, Kenneth M. (1985). Applying sociology to policy: Social science and the environmental impact statement. *Rural Sociology*, 50, 578-604.

Green, Charles S., III, & Salem, Richard G. (1982). The nonsociologist as applied sociologist: Teaching undergraduates applied sociology as ethical practice. *Teaching Sociology*, 11, 32-46.

Johnson, Doyle Paul, Brown, William R., Hage, Jerald, Lyson, Thomas A., Orthner, Dennis K., Paulson, Steven K., Squires, Gregory D., & Wimberley, Ronald D. (1987). The challenge of training in applied sociology. *American Sociologist*, 18, 356-368.

Kelly, Robert F. (1986). Teaching graduate applied sociology through internships: Program development, management, and problems. *Teaching Sociology*, 14, 234-242.

Kenig, Sylvia. (1992). *Who plays? Who pays? Who cares? A case study in applied sociolo-*

gy, political economy, and the community mental health centers movement. Amityville, NY: Baywood.

Kimmel, Allan J. (1988). *Ethics and values in applied social research* (Vol. 12). Thousand Oaks, CA: Sage.

Lazarsfeld, Paul, & Reitz, Jeffery G. (In press). *An introduction to applied sociology.* Ann Arbor, MI: Books on Demand.

Maneker, Jerry S. (1994). *Applied sociology: Sociological understanding and its application.* Lanham, MD: University Press of America.

Olsen, Marvin E., & Micklin, Michael. (Eds.). (1981). *Handbook of applied sociology: Frontiers of contemporary research.* New York: Praeger.

Park, Betty Jean. (1982). The Center for the Study of Local Issues: A vehicle for applied sociology. *Teaching Sociology*, 11, 105-112.

Parsons, Talcott. (1997). *Essays on sociological theory: Pure and applied.* New York: Free Press.

Petrus, Gene, & Adamek, Raymond J. (1988). Taking the role of the other: An aid to marketing applied sociology. *Teaching Sociology*, 16, 25-32.

Popenoe, David. (2000). *Sociology.* Upper Saddle River, NJ: Prentice Hall.

Porter, M. (1998). *Psychology and sociology applied to medicine.* Camden Town, UK: Churchill-Livingstone.

Sherohman, James. (1982, August-September). *Applied sociology and social work.* Paper presented at the annual meeting of the American Sociological Association, Detroit, MI.

Steele, Stephen F., & Iutcovich, Joyce M. (Eds.). (1997). *Directions in applied sociology: Presidential addresses of the Society for Applied Sociology*, 1985-1995. London: Arnold.

Steele, Stephen F., Scarsisbrick-Hauser, Annemarie, & Hauser, William J. (1998). *Solution-centered sociology: Addressing problems through applied sociology.* Thousand Oaks, CA: Sage.

Sullivan, Thomas. (1991). *Applied sociology: Research and critical thinking.* Upper Saddle River, NJ: Prentice-Hall.

Tepperman, Lorne, & Blaine, Jenny. (1999). *Think twice! Sociology looks at current social issues.* Upper Saddle River, NJ: Prentice-Hall.

Wallace, Richard Cheever. (1988). A capstone course in applied sociology. *Teaching Sociology*, 16, 34-40.

Ward, Lester F. (1999). *Applied sociology.* Temecula, CA: Reprint Services Corporation.

Watts, W. David, Short, A., & Schultz, C. (1982). Applied sociology and the current crisis. *Teaching Sociology*, 11, 47-61.

Wright, R., Dubois, William, & Bacon, Allyn. (2000). *Applying sociology: Making a*

better world. Needham Heights, MA: Allyn and Bacon.

Yiannakis, Andrew, & Greendorfer, Susan L. (Eds.). (1992). *Applied sociology of sport.* Champaign, IL: Human Kinetics.

3.6.2 知识应用

Backer, Thomas E., & Shaperman, Julie. (1992). Knowledge utilization and foundations supporting health research and demonstrations: Initial explorations. *Knowledge: Creation, Diffusion, and Utilization*, 14, 386-400.

Beker, Jerome, & Eisikovits, Zvi C. (Eds.). (1991). *Knowledge utilization in residential care and youth care practice.* Washington, DC: Child Welfare League of America, Inc.

Cernada, George P. (1982). *Knowledge in action: A guide to research utilization.* Amityville, NY: Baywood.

Guiltinan, Joseph P., & Achabal, Dale. (1986). *AMA Winter Educators' Conference, 1986: Marketing Education: Knowledge Development, Dissemination, and Utilization.* Ann Arbor, MI: Books on Demand.

Havelock, Ronald G., & Guskin, Alan E. (1971). *Planning for Innovation Through Dissemination and Utilization of Knowledge.* Ann Arbor, MI: Center for Research on Utilization of Scientific Knowledge.

Hess, T. M. (Ed.). (1990). *Aging and cognition: Knowledge organization and utilization.* New York: Elsevier Science.

Hultman, Glenn, & Horberg, Cristina Robertson. (1998). Knowledge competition and personal ambition: A theoretical framework for knowledge utilization and action in context. *Science Communication*, 19, 328-348.

Kiresuk, Thomas J. (1992). The evaluation of knowledge utilization: Placebo and nonspecific effects, dynamical systems, and chaos theory. *Journal of the American Society for Information Science*, 44, 235-241.

Paisley, William. (1992). Knowledge utilization: The role of new communication technologies. *Journal for the American Society of Information Science*, 44, 222-234.

Wong, Kenneth K. (1998). Laying the groundwork for a new generation of policy research: Commentary on "Knowledge utilization in educational policy and politics," Spotlight on Student Success No. 302. *Educational Administration Quarterly*, 34, 141-146.

3.6.3 政策制订/政策分析

Anderson, James E. (2000). *Public policymaking: An introduction* (4th ed.). Boston: Houghton Mifflin.

Bardach, Eugene. (2000). *A practical guide for policy analysis: The eigbtfold path to more effective problem solving.* New York: Seven Bridges Press.

Berman, David R. (1982). *American government, politics, and policymaking.* Pacific Palisades, CA: Palisades Publishers.

Burns, Tom R., & Ueberhorst, Reinhard. (1988). *Creative democracy: Systemic conflict resolution and policymaking in a world of high science and technology.* Westport, CT: Greenwood.

De-Greene, Kenyon B. (Ed.). (1992). *A systems based approach to policymaking.* Norwood, NJ: Kluwer Academic Publishers.

Friedelbaum, Stanley. (1988). *Human rights in the states: New directions in constitutional policymaking.* Westport, CT: Greenwood.

Janis, Irving L. (1989). *Crucial decisions: Leadership in policymaking and crisis management.* New York: Free Press.

Longest, Beaufort B. (1998). *Health policymaking in the United States.* Chicago: Health Administration Press.

Meehan, Eugene J. (1990). *Ethics for policymaking: A methodological analysis.* Westport, CT: Greenwood.

Mezey, Susan G. (1996). *Children in court: Public policymaking and federal court decisions.* Albany: State University of New York Press.

Protess, D., & McCombs, M. (Eds.). (1991). *Agenda setting: Readings on media, public opinion and policymaking.* Mahwah, NJ: Lawrence Erlbaum.

Prunty, John J. (1984). *A critical reformulation of educational policy analysis.* New York: State Mutual Book & Periodical Service.

Rich, Michael J. (1992). *Federal policymaking and the poor: National goals, local choices, and distributional outcomes.* Princeton, NJ: Princeton University Press.

Ross, Robert S. (Ed.). (1992). *China, the United States and the Soviet Union: Tripolarity and policymaking in the Cold War.* Armonk, NY: M. E. Sharpe.

Schmitz, David F., & Jespersen, T. Christopher. (Eds.). (1999). *Architects of the American century: Individuals and institutions in twentieth-century U. S. foreign policymaking.* Chicago: Imprint Publications.

Sharkansky, Ira. (1999). *Ambiguity, coping and governance: Israeli experiences in politics, religion and policymaking.* Westport, CT: Greenwood.

Shepard, N. Alan. (1984). *A policy analysis of professional development and personnel preparation for serving special populations.* Columbus, OH: Center for Education and Training for Employment.

Stokey, Edith, & Zeckhauscr, Richard. (1978). *A primer for policy analysis.* New York: W. W. Norton and Company.

3.6.4 评 价

Ackerman, Marc J. (1995). *Clinician's guide to child custody evaluations.* New York:

John Wiley and Sons.

Alkin, Marvin C., & Jacobsen, Phyllis. (1985). *A guide for evaluation decision makers.* Lima, OH: Books on Demand.

Boruch, Robert F. (1997). *Randomized experiments for planning and evaluation: A practical guide.* Thousand Oaks, CA: Sage.

Davitz, Joel Robert, & Davitz, Lois Leiderman. (1996). *Evaluating research proposals: A guide for the behavioral sciences.* Englewood Cliffs, NJ: Prentice Hall.

Fink, Arlene. (1992). *Evaluation fundamentals: Guiding health programs, research and policy.* Newbury Park, CA: Sage.

Giardino, Angelo P., Christian, Cindy W, & Giardino, Eileen R. (1997). *A practical-guide to the evaluation of child physical abuse and neglect.* Thousand Oaks, CA: Sage.

Giardino, Angelo P., Finkel, Martin A., Giardino, Eileen R., Seidl, Toni, & Ludwig, Stephen. (1992). *A practical guide to the evaluation of sexual abuse in the prepubertal child.* Newbury Park, CA: Sage.

Korte, Nic. (1999). *A guide for the technical evaluation of environmental data.* Lancaster, PA: Technomic.

Ogles, Benjamin M., Lambert, Michael J., & Masters, Kevin S. (1996). *Assessing outcome in clinical practice.* Boston: Allyn and Bacon.

Patrick, John. (1992). *Training: Research and practice.* London: Academic Press.

Patton, Michael Quinn. (1997). *Utilization-focused evaluation: The new century text.* Thousand Oaks, CA: Sage.

Rice, Marilyn, & Valdivia, Leonel. (1991). A simple guide for design, use and evaluation of educational materials. *Health Education Quarterly*, 18, 79-85.

Rubin, Frances. (1995). *A basic guide to evaluation for development workers.* Oxford, UK: Oxfam Publishing.

Simon, Robert I. (Ed.). (1995). *Posttraumatic stress disorder in litigation: Guidelines for forensic assessment.* Washington, DC: American Psychiatric Press.

Stahl, Philip Michael. (1994). *Conducting child custody evaluations: A comprehensive guide.* Thousand Oaks, CA: Sage.

3.6.5 评价的范例

Al-Sharideh, Khalid. (1999). *Modernization and socio-cultural transformation in Saudi Arabia: An evaluation.* Unpublished doctoral dissertation, Kansas State University.

Altus, Deborah E., & Mathews, R. Mark. (1999). A case study evaluation of the home-care suite: A new long-term care option for elders. *Journal of Housing for the Elderly*, 13, 115-125.

Arts, Wil, Hermkens, Piet, & Van-Wijck, Peter. (1999). Modernisation theory, income evaluation, and the transition in Eastern Europe. *International Journal of Comparative*

Sociology, 40, 61-78.

Atim, Chris. (1999). Social movements and health insurance: A critical evaluation of voluntary, non-profit insurance schemes with case studies from Ghana and Cameroon. *Social Science and Medicine*, 48, 881-896.

Berk, Richard A., & de-Leeuw, Jan. (1999). An evaluation of California's inmate classification system using a generalized regression discontinuity design. *Journal of the American Statistical Association*, 94, 1045-1052.

Birnbacher, Dieter. (1999). Quality of life—Evaluation or description? *Ethical Theory and Moral Practice*, 2, 25-36.

Bode, Maarten. (1998). Social structure and change: Vol. 1. Theory and method. An evaluation of the work of M. N. Srinivas. *Contributions to Indian Sociology* (New Series), 32,558-559.

Bogner, Franz X., & Wiseman, Michael. (1998). Environmental perception of Swiss and Bavarian pupils: An empirical evaluation. *Revue Suisse de Sociologie*, 24, 547-566.

Bonta, James, Wallace, Capretta-Suzanne, & Rooney, Jennifer. (2000). Can electronic monitoring make a difference? An evaluation of three Canadian programs. *Crime and Delinquency*, 46, 61-75.

Calloway, Michael O., & Nadlicki, Terri M. (2000, April). *Research and learning: Using theories of social capital to guide health policy and evaluation.* Paper presented at the annual meeting of the Southern Sociological Society, New Orleans, LA.

Cannon, Julie Ann Harms. (1996). The social construction of knowledge: An evaluation of the works of Dorothy E. Smith and Patricia Hill Collins. *Humanity and Society*, 20, 9-24.

Chodos, Howard. (1998). Theory and metatheory in the evaluation of Marxism: A review essay. *Critical Sociology*, 24, 156-166.

Craglia, Massimo, Haining, Robert, & Wiles, Paul. (2000). A comparative evaluation of approaches to urban crime pattern analysis. *Urban Studies*, 37, 711-729.

Dalenberg, Constance J. (1999). Overcoming obstacles to just evaluation and successful prosecution of multivictim cases. *Journal of Aggression Maltreatment and Trauma*, 2, 141-162.

Dumaine, Marian Lee. (1998). *An evaluation of coping ability as a guide to the treatment of persons diagnosed with schizophrenia and substance abuse.* Unpublished doctoral dissertation, Florida International University, Miami.

Farrell, Graham, & Buckley, Alistair. (1999). Evaluation of a UK police domestic violence unit using repeat victimisation as a performance indicator. *Howard Journal of Criminal Justice*, 38, 42-52.

Ferligoj, Anuska, & Hlebec, Valentina. (1999). Evaluation of social network measurement instruments. *Social Networks*, 21, 111-130.

Foddy, William. (1998). An empirical evaluation of in-depth probes used to pretest survey questions. *Sociological Methods and Research*, 27, 103-132.

Gavazzi, Stephen M., Wasserman, Deborah, Partridge, Charles, & Sheridan, S. (2000). The Growing Up FAST diversion program: An example of juvenile justice program development for outcome evaluation. *Aggression and Violent Behavior*, 5, 159-175.

Gransky, Laura A., & Patterson, Marisa E. (1999). A discussion of Illinois' gang free prison: Evaluation results. *Corrections Management Quarterly*, 3, 30-42.

Green Mazerolle, Lorraine, Ready, Justin, Terrill, William, & Waring, Elin. (2000). Problem-oriented policing in public housing: The Jersey City evaluation. *Justice Quarterly*, 17, 129-158.

Harris, Philip W., & Jones, Peter R. (1999). Differentiating delinquent youths for program planning and evaluation. *Criminal Justice and Behavior*, 26, 403-434.

Hatley, Heather Lynne. (1999). *The influences of managed care on certified nurse midwives: An evaluation of health system change.* Unpublished doctoral dissertation, University of Wisconsin, Madison.

Henderson, Harold, German, Victor F., Panter, A. T., Huba, G. J., Rohweder, Catherine, Zalumas, Jacqueline, Wolfe, Leslie, Uldall, Karina K., Lalonde, Bernadette, Henderson, Ron, Driscoll, Mary, Martin, Sara, Duggan, Sandra, Rahiman, Afsaneh, & Melchior, Lisa A. (1999). Systems change resulting from HIV/AIDS education and training: A cross-cutting evaluation of nine innovative projects. *Evaluation and the Health Professions*, 22,405-426.

Henry, Gary T., & Julnes, George. (1998). Values and realist evaluation. *New Directions for Evaluation*, 78, 53-71.

Holmila, Marja. (1999). Public policy and program evaluation. *Acta Sociologica*, 42, 97-99.

Jasso, Guillermina, & Wegener, Bernd. (1999). Gender and country differences in the sense of justice: Justice evaluation, gender earnings gap, and earnings functions in thirteen countries. *International Journal of Comparative Sociology*, 40, 94-116.

Julnes, George, & Mark, Melvin M. (1998). Evaluation as sensemaking: Knowledge construction in a realist world. *New Directions for Evaluation*, 78, 33-52.

Khan, M.-Adil. (1998). Evaluation capacity building: An overview of current status, issues, and options. *Evaluation*, 4, 310-328.

Klein, Susan Shurberg. (1992). Sharing the best: Finding better ways for the federal government to use evaluation to guide the dissemination of promising and exemplary education solutions. *Evaluation and Program Planning*, 16, 213-217.

Koskinen, Ilpo. (2000). Plans, evaluation, and accountability at the workplace. *Sociological Research Online.* Retrieved February 4, 2000, from the World Wide Web: http://www. socresonline. org. uk

Kuehne, Valerie-Shahariw. (1999). Building intergenerational communities through research and evaluation. *Generations*, 22, 82-87.

Kyudd, Sally Anna. (1999). *A case study of program planning and evaluation in assisting Montserratian evacuees and British government officials in natural disaster planning.* Unpublished doctoral dissertation, Rutgers State University.

Larkin, Carl A., & Knowlton, Amy R. (2000). New directions in HIV prevention among drug users: Settings, norms, and network approaches to AIDS prevention (SNNAAP); a social influence approach. *Advances in Medical Sociology*, 7, 261-287.

Leon, Ana M., Dziegielewski, Sophia F., & Tubiak, Christine. (1999). A program evaluation of a juvenile halfway house: Considerations for strengthening program components. *Program Planning and Evaluation*, 22, 141-152.

Loveday, Barry. (1999). Tough on crime or tough on the causes of crime? An evaluation of Labour's crime and disorder legislation. *Crime Prevention and Community Safety—An International Journal*, 1, 7-24.

Mason, George P. (1999). *The Congress of Industrial Organizations and the rise of American industrial unionism: Are-evaluation of the record.* Paper presented at the annual meeting of the American Sociological Association.

Mears, Daniel P. (1998). Evaluation issues confronting juvenile justice sentencing reforms: A case study of Texas. *Crime and Delinquency*, 44, 443-462.

Mears, Daniel Preston. (2000). *Evaluating juvenile justice reform: An analysis of prosecutorial discretion and determinate sentencing in Texas.* Unpublished doctoral dissertation, University of Texas, Austin.

Miller, J. Mitchell. (1999). Administrative police training for an emerging democracy: An evaluation of the Moscow Police Command College—a research note. *American Journal of Criminal Justice*, 23,267-276.

Moore, Melanie, Blumstein, Philip, & Schwarz, Pepper. (1998). The power of motherhood: A contextual evaluation of family resources. *Free Inquiry in Creative Sociology*, 26, 111-117.

Neuhauser, Linda, Schwab, Michael, Syme, S. Leonard, Bieber, Michelle, & Obarski, Susan King. (1998). Community participation in health promotion: Evaluation of California wellness guide. *Health Promotion International*, 13, 211-222.

O'Sullivan, Rita G., & O'Sullivan, John M. (1998). Evaluation voices: Promoting evaluation from within programs through collaboration. *Evaluation and Program Planning*, 21, 21-29.

Rossi, Peter H. (1997). Advances in quantitative evaluation 1987-1996. *New Directions for Evaluation*, 76, 57-68.

Rowe, Gene, & Frewer, Lynn J. (2000). Public participation methods: A framework for evaluation. *Science, Technology, and Human Values*, 25, 3-29.

Shapiro, Jeremy P., Welker, Carolyn J., & Pierce, Janice L. (1999). An evaluation of residential treatment for youth with mental health and delinquency-related problems. *Residential Treatment for Children and Youth*, 17, 33-48.

Slizyk, Jeanne. (2000, April). *Program evaluation of a program to assist victims of domestic violence.* Paper presented at the annual meeting of the Southern Sociological Society, New Orleans, LA.

Stall, Ron D., Paul, Jay B., Barrett, Donald C., Crosby, G. Michael, & Bein, Edward. (1999). An outcome evaluation to measure changes in sexual risk taking among gay men undergoing substance abuse disorder treatment. *Journal of Studies on Alcohol*, 60, 837-845.

Stockmann, Reinhard. (1999). The implementation of dual vocational training structures in developing countries: An evaluation of "dual projects" assisted by the German Agency for Technical Cooperation. *International Journal of Sociology*, 29, 29-65.

Van Goor, H., & Stuiver, B. (1998). Can weighting compensate for nonresponse bias in a dependent variable? An evaluation of weighting methods to correct for substantive bias in a mail survey among Dutch municipalities. *Social Science Research*, 27, 481-499.

Van Hees, Alma. (1999). Halt: Early prevention and repression; recent developments and research. *European Journal on Criminal Policy and Research*, 7, 405-416.

Voas, R. B., Marques, P. R., Tippetts, A. S., & Beirness, D. J. (1999). The Alberta Interlock Program: The evaluation of a province-wide program on DUI recidivism. *Addiction*, 94, 1849-1859.

Wagner, Peter. (1999). After justification: Repertoires of evaluation and the sociology of modernity. *European Journal of Social Theory*, 2, 341-357.

Williams, James L., Rodeheaver, Daniel G., & Huggins, Denise W. (1999). A comparative evaluation of a new generation jail. *American Journal of Criminal Justice*, 23, 223-246.

Wilson, Barbara Renee. (1999). *Evaluation of the Family Support Worker Program.* Unpublished doctoral dissertation, University of Minnesota.

Wilson, Steve T. (2000). *Minority police officers, police discrimination and the police subculture: An evaluation of prior studies and a proposal for future research.* Paper presented at the annual meeting of the Southern Sociological Society.

Yeboah, David. (2000). The evaluation of New Zealand's Habilitation Centre's pilot program. *Journal of Criminal Justice*, 28, 227-235.

3.6.6 知识应用的范例

Berlin, Donna F., & White, Arthur L. (1992, April). *Action research as a solution to the problem of knowledge utilization.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Cousins, J. Bradley, & Leithwood, Kenneth A. (1992). Enhancing knowledge utilization as a strategy for school improvement. *Knowledge: Creation, Diffusion, and Utilization*, 14, 305-332.

Datta, Lois Ellin. (1992). A grass roots perspective on legislating knowledge utilization: The Seventh Annual Howard Davis Memorial Lecture, April 1992. *Knowledge: Creation, Diffusion, and Utilization*, 14, 291-304.

Elpers, J. R. (1989). Development and utilization of incentive systems for mental health operations: Successful and unsuccessful knowledge utilization in California and Los Angeles. *Journal of Mental Health Administration*, 16, 9-20.

Nelson, Scott J. (1989). Values and priorities: Their effect on knowledge utilization in public mental health programs. *Journal of Mental Health Administration*, 16, 44-49.

Peterson, Paul D., & Quasim, Lyle. (1989). Using strategy to promote knowledge utilization and change in community mental health systems. *Journal of Mental Health Administration*, 16, 29-36.

Sorensen, James E. (1989). Collaboration among state and local mental health organizations: Examples of knowledge utilization. *Journal of Mental Health Administration*, 16, 3-8.

评价研究的过程 3.7

任何试图减少或根除某个社会问题的努力都将涉及一种理论、一个项目以及经常是很大的一笔资金。那些关于减少犯罪和不良行为、阻止毒品上瘾、克服健康问题以及改善邻居、群体关系和改善生活质量方面的项目效果都引发了很多评价问题。因为这些问题对国家的和群体的生活非常重要,并且它们耗费巨大,因此,对它们的评价被给予了一个很高的优先地位,评价研究也逐渐增多起来。

Edward Suchman(1967)曾经这样写道:

> 评价常常开始于某些价值观念(value),或者是内在的,或者是外在的。比如说,"活得时间长是好事"是一种关于价值判断的观念,而从这种观念中就产生了一个目标。因此,目标的选择在"观念形成"之后或与之同时进行。举个"目标设定"(goal setting)的例子,比如:更少的人将会得冠心病,或者说,不应有这么多的人会死于癌症。目标设定相互之间竞争的动力来源于资金、资源和成就。
>
> 接下来的过程是"测量目标的结果"。假如我们设定我们的目标是较少的人会死于癌症,于是我们就需要一些方法去发现目前死于癌症的有多少人(比如,死亡率的统计)。评价的种类在很大程度上取决于我们可以获得并

可用来决定我们目标达到的方法。在这个过程中接下来的一步是明确“达到目标的行动”类型。就癌症的例子来说，一个致力于尽早发现和治疗癌症的行动项目是应该被考虑的。接着，这种行动付诸实施。诊断中心因此建立，人们也因此被督促来这里检查。

于是，在某一点上，我们对这种目标指向的操作有了评定。这一阶段包括了对这个操作运行后所达到的相对于事先决定目标的程度评价。正如刚才谈到的，这种评定可能做得很科学，但也可能做得不科学。

最后，在这个评定的基础上，这种目标指导行为是否值得也就有了一个判断。它使我们重新回到评价组合的问题上。此时有些人可能会说有癌症诊断中心是很好的。在评价过程的最后，我们可能会得到一种新的评价，或者我们会重新再肯定，再评价，或者重新定义一种旧的评价。比如，假如旧的理念是“活得时间长是好事”，新的理念可能是“你健康地活到100岁是很好的事情，但是如果你不能保持健康，别活过80岁更好”。

参考文献

Suchman, Edward A. (1967). *Evaluative research, principles and practice in public service and social action programs.* New York: Russell Sage Foundation.

3.7.1 评价模型的设计和应用

许多社会的和行为学科发生了改变。这种改变方式表现为一种逐渐加大的对可说明性的呼吁。这些方式中的一种已经被表述过，它贯穿了评价模型和策略的设计及实施全过程。在接下来摘录中，Rossi 和 Wright 讨论了评价在项目后果评定中的角色，以及当评价作为一种科学的方法去回答某个项目是否有效时，它如何发展。

评价(evaluation)研究：一种评估(assessment)

Peter H. Rossi & James D. Wright

20世纪60年代中期的“大社会项目”(great social programs)时期，评价研究作为应用社会科学地位变得非常突出。过去25年最突出的特征是政策制订者和公共管理者有一个非常清楚的认识：评价可以被系统地引入到社会科学研究方法的使用中去，并且较之早先使用的审判方法可以产生更有用和更正确的结论。“大社会项目”时期，国会在一些非常重要的立法领域批准了许多新的项目，系统性评价在几个非常重要的领域也被使用。[1]

那些成立起来为完成这些项目的新的管理代理机构里，有部分人员是社会科学家，这些人在应用工作上有非常浓厚的兴趣。社会科学的整个领域都被包括进来。在经济发展机会署，经济学家有很重要的地位；社会学家、心理学家以及教育家也在教育署(后称为教育部)有很重要的地位；健康部、教育部、福利部(现统称为健康与人类服务部)在非常重要的位置上给社会科学人员提供了足够的发展机会；同时，劳动人力

资源研究分部也非常慷慨,给这些人提供了许多机会。

这种新的社会科学行为所具有的交叉学科的特征非常值得提及。经济学家、社会学家、心理学家以及教育研究者经常发现彼此在为同一个契约而竞争,这样的过程有利于跨学科的知识、技能的交流,增进彼此的尊敬。早先局限于某一个领域的研究公司和机构为了增加自身的竞争力,雇佣了其他社会科学的专业人士而拓宽了自身的发展前景。跨学科的专业性群体也建立起来,比如,"评价研究会和评价网络"的建立。[2]

以社会科学研究者为主的大学利用这种新的机会去获取研究基金的速度很慢,即使涉及的论题往往是很重要的。这种现象折射出一种对应用性工作的冷漠(甚至是敌视),即使在不久前,它都一直在各种相关的理论社会科学系里存在(Raizen & Rossi 1981,Rossi & Wright 1983, Rossiet al . 1978)。然而,私人企业,却非常迅速地注意到把评价作为新的重点来开发利用。现在一些对社会科学并不特别感兴趣的公司开的一些子公司却在社会研究上具有竞争力(如,Westinghouse)。其他的公司也在迅速地扩充他们的社会研究部门(如,Rand 公司)。此外,逐渐地,成百的公司呈现出这种景象,在这段黄金时期,其中有一部分还特别成功(如,Abt 协会)。[3]

到20世纪70年代中期,大约500~600家私人公司主要从事于应用社会研究。但正如在其他领域的公司活动一样,少数公司获得了大部分可利用的资金。举个例子,在1975—1980年间,6家大的研究公司得到了大约来自教育部资金的60%(Raizen & Rossi 1981)。

在州政府和地区层次,出现了一批其他类型的公司,主要从事于测评和应用社会研究活动。这些研究的机遇,既不像那些由联邦政府资助的机构那样有好的资金支持,在任务分派的机会上也不如智力的或学术上的机构那么好。但是,在州政府和地区层次已经并持续会有足够的评价商机,给很大一部分小规模的商家提供了非常重要的"面包和黄油"。

在这期间,一些有着丰富历史性的大规模社会研究的大学研究机构也处于繁荣阶段。芝加哥大学的全国民意研究中心和密歇根大学的调查研究中心都发展到特别大的规模。新的学术研究机构也开始利用通过契约机制所给予的在基金方面的机遇。

一个紧随而来的进步发生在评价研究的概念方面。1966年,从事研究设计基础工作(这些工作对教育评价项目非常有用)的 Donald t. 和 Julain Stanley 有了出版物,这个出版物为研究设计的分类和讨论论题的有效性创造了一个完整的新词汇。它同样也创造了一种随机的、控制性的实验典范,为因果研究提供了选择方法。在接下来的十年里,这两大进步统治了评价研究的大部分领域。最初,评价研究主要是用来评价整个项目网络的效果。慢慢地,设计评价研究时的主要问题是去详细说明适当的标准条件,这可以保证对这些网络效果正确有效的估计。在这整个框架里,随机、控制性实验成为评价研究的统治性典范。这种概念性的基础在几十年前就已经发展起来,而这种方法在心理学和生物学上成为统治性范例也已有很多年。通过回顾,可以发现这段时间特殊的贡献是这种范例来自于实验室,应用于研究领域,并且它和在研究中设计用来测试被计划的项目效果的案例研究联系起来。这对倾向于使用技术手段的社会科学家来说,随机化的实地实验可以确保社会政策拥有一种更具有自我意识的理性

的“实验社会”,从而替代我们现在拙劣的试错法(Campbell,1969)。

到20世纪70年代早期,数量多的惊人的大规模实验建立并开始实施。这些实验覆盖了非常广泛的主题:意图用来替代现在的福利系统的收入维持计划,刺激市场产生更多房屋给穷人的房屋津贴,不会产生不正当医疗护理价格结果的健康保险计划。诸如此类的实验通过一系列实验细目清单产生。具有讽刺意味的是,这些实验的大部分都是由经济学家来设计和实施,这些领域的成员中没有人提到这些实验工作最初的传统形态。

不过,人们很快意识到,随机的、受控制的实验只能在有限的环境控制下操作才能正确,并且许多简单项目的评价需求在这里不能实现。不仅随机性所需求的一些情况频繁地被道德和法律所限制,而且许多现有的完全(或几乎)覆盖了项目自身意图的利益群体的项目无法使用控制组来评价,因为没有任何方法去创造一个非常精确的控制组。同时,这种研究也显得非常耗时,从开始设计到最终报告要花上3~5年,甚至更长的时间,对于只有简单容忍力并需要更短时间的大多数政策制订者和公众管理者来说,这是一种耽搁。

Campbell和Stanley(1966)提供了一种可行的方法去解决这种困境。这种方法是通过整合准实验的方式,并使用这种方式去代替评价研究设计,不用依赖随机就可得到控制组。尽管他们清楚地意识到通过这种方法得到的数据有效性稍差,但他们同样也讨论了,通过在这种方法下使用评价研究可以得到正确因果关系的推论。他们对准实验研究的设计处理的确刺激了评价研究中对这种设计的使用,一般是在Campbell和Stanley清楚陈述过的一些非常重要的条件下进行的。事实上,绝大多数进行的评价都是准实验型的,而不是随机的、“真正的”实验,这主要是因为后者被证明要用现实世界的工具来实现太难了,几乎不可能。

但是即使是准实验也是有缺陷的。首先,尽管不像“真正的”实验那么昂贵、耗时,一个运作很好的准实验可能要求更多超出可获得的资金、时间和忍耐力。另一个问题是许多更复杂的准实验设计(特别是,断续时间系列设计)要求更长时间序列的数据——理想上,这种序列包含一个长时期的观察,这种观察在一个政策要实现之前,以及在政策实施后的几年内持续存在。可以肯定的是,这些必需的数据事实上往往不存在。同时,可以肯定的第二点是,一些老问题在适当的时候反复出现。最后一个问题,Campbell和Stanley非常详细地讨论过:存在潜在的对准实验设计的有效性有威胁的因素。使用这种设计,想得到真正项目效果,会更多地经历得到各种虚假结果的危险。因此,准实验经常容易受到批评性的攻击。一些主要的教育性评价便经历了许多激烈的争论(比如,McLaughlin 1975, Mosteller & Moynihan 1972, Rossi & Wright 1982)。

由于实验的和准实验的方法用来测评时都有明显的问题,对于许多害怕评价会在某种程度上将其套住的项目管理者和官员而言,显然非常需要及时的、相对便宜的、反应性的评价方法。尤其,对那些由国会立案的和项目代理希望自己能管理的评价,这种担心就更多。事实上,国会——把它新的对评价的热情和某个严重错误的理解联系起来,这种理解是对时间、潜力以及对需要完成起码质量的评价所需资金的理解——经常把这种评价任务强加给项目代理,但是这已经超出了代理们的研究能力和拥有的

资金能力,因为他们的资金不足以应付他们要做的事情。

一些评价可以由一些技术好而并非资历深的人来完成,并且这些评价对项目的管理者们非常及时和有用,而存在的对这些评价的需要就激发了对用于评价研究的定性研究法的巨大兴趣(Patton 1980,Scriven 1977,Guba & Lincoln 1981, House 1980)。定性研究方法一直以来在各种社会科学中就或多或少存在,尤其是社会学。在社会学中,它们特殊的吸引力来自于:通过对研究领域情况的了解,它们被假定认为具有贴近实际以及可以促进理解社会的进程的能力。此外,对于评价目的而言,定性研究看起来有三个非常重要的优点:耗费不大,及时,以及对领导者的需求具有反应力。这些方面对于项目的资助者和实行者具有非常大的吸引力,因为它们呈现出来的是具有足够的弹性去应付社会项目,一旦完成,它所发挥出来的作用涉及面非常广,从一个领域到另一个,不仅是它们最初的目标,只要是它们波及的都将受益。一些跨及多领域的项目目标(比如,样板城市)并没有非常清晰地被国会或者是管理机构确定。因此每一个实施机构就必须明确它们自己的目标并且经常性地改变(Kaplan 1973, Williams 1980)。这种对定性研究评价的诉求,至少在最开始的时候,表现为它们明显地对少许的错误定义具有敏感性以及经常地改进自己的项目目标。

在 1981 年,评价拥有的这种盛事结束了。这个时候里根政府开始减少那些在过去的 20 年里发展非常好的社会项目。劳动部有关广义人力资源的研究项目被减少到几乎没有,这种相同的境遇(虽然没有这么夸张)也同样发生在健康与人类服务部、教育部以及农业部等等其他部。紧随而来的是用于应用研究方面的州立资金急剧减少。

具有讽刺意味的是,里根政府的这种削减措施正好伴随着越来越多的学术部门发现非学术市场提供了机会给新出现的博士生们。公开评价研究是这个市场最大的组成元素。1981 年,在华盛顿美国社会学会召开了有许多人参加的会议,讨论建立一套适当的培训方法去适应应用社会学方面的职业。在全国各地有很多研究生院,计划开设培训各种各样的研究者,很多杰出的社会学家在这方面表现了很大的兴趣。事实上,美国社会学协会的两个主席在 1980 年和 1981 年都把他们主要的事务放在应用工作上(Rossi 1981,Whyte 1982)。

评价黄金期的人力丰收

处于 20 世纪 60 年代和 70 年代,评价研究的狂热发展促进了我们有关社会问题知识的切实增长,并且极大促进了有关社会科学技术的深化。这两个发展已经在社会科学上产生了一些影响,将来对我们的研究领域也会越来越有价值。

"黄金期"的大规模实地实验

整个黄金期给人印象最深的切实性和技术性的成就大概是大规模实地实验。这种实验最开始大多是由经济发展机会部门计划的,后来由于这个部门的消失,就由健康、教育和福利部门来做这件事。

在技术方面,这些实验联结了样本调查技术(sample survey techniques)和经典实验设计(classical experimental designs)。实验组和控制组都取样于开放的社区,然后随机地分派样本家庭给实验组和控制组。对实验组及控制组的家庭访问开始运作,并使

用传统样本调查的技术去测量实验性对待的反应。作为调查,这些实验有长时期的工作小组来反复测量主要的自变量(比如:结果)。在某些实验里测量几乎是一个月一次,延时将近五年。作为实验,研究是因子式的,在它里面的重要参数处理总在系统地改变。

在黄金期最出名的实地实验是那些设计用来测量各种形式的"不良收入税"(NIT),这种税收为贫困家庭维持了合理的最低收入保障线。总地来说,在美国有五个这样的实验,在加拿大有一个。

修补非实验设计

到目前为止,讨论已经非常窄地聚集在有关影响评价的随机实验设计上,因为影响评价技术性地成功是通过使用这种设计实现的,因此那些随机的、控制性的实验控制着整个评价体系至少达两个世纪。然而,即使有上面的这些细节,这里仍有很多好的理由要去修正实验范式,主要是因为对于许多社会项目必须使用非实验的方法。

这里有很多原因来解释为什么随机实验设计不能被用于某些评价研究中。首先,那些正在进行的、覆盖了大部分或者它们希望的全部目标人群的项目不一定是受到了可以信任的控制。比如,估计5%~10%符合老年生存保险(社会保险)福利的人并没有享受到这种应有的待遇。然而,实际上这些作为控制组用来测试社会福利效果的非申请者并不能起到相应的作用,因为自选因素毫无疑问是最强的。把那些接受到社会保险利益的人和不管任何原因都具有资格却没有享受到的人相比,违反了标准无差别的准则(ceteris paribus condition)。

第二,一些项目,比如,重点启蒙项目,无法得到合适比例的够资格的人群——可能只达到了需要的25%。有些孩子没有进入重点启蒙项目,是因为他们的父母不愿意自己的孩子加入,或者是因为与此有关的学校没有这么多的穷孩子来支持这个项目。很明显,自选因素起了作用,因此,对比重点启蒙项目的参加者和够资格的非申请者,这两者之间不会存在经常性严重的差异。

最后,伦理上认为,在某些项目评价上使用随机是无法想像的。比如,一个确定用来估算私立学校和公立学校相关效力的方法,可能随机地指派青少年在某一组,并在很长一段时间里观察(在他们身上)可能(出现)的结果。很明显,没有任何情况,或者是政策制订者或者父母们会愿意同意这种评价发生。

因此,在过去的二十多年里,除经典随机实验设计,评价研究运用了一些其他的方法。不幸的是,在总体上,这些评价并没有获得技术上的胜利。每个主要的非实验评价在争论中被掩盖,这种争论产生于研究发现的政治性应用之外,而常常主要集中在所使用的设计的技术缺乏上。因此,Coleman(1966)通过对从千百所中学挑出的千余名学生的跨学科分析,试图整理出学校对成绩产生影响的行为方式受到了批评,批评主要是针对他运用的统计模型(Mosteller & Mounihan 1972)。同样,一个关于参加重点启蒙项目的长期效应评价(Westinghouse Learning Corporation 1969)也遭到了批评,原因在于研究者把那些参加重点启蒙项目的青少年和那些条件也合适但却没有参加的青少年做了比较。按照批评者们的说法,忽视自选因素的影响将导致两个组群在很多重要的方面都不可比。

与管理性或者自选性项目参与者和非参与者相关的问题是:非实验评价设计的核心容易受到批评。为了更好地阐释这一点,我们可以来看看近来 Coleman 以及他的助手在公立学校和私立学校的学术成就比较方面的研究。这个研究中评论性的比较就受到了自选因素的负面影响:一个孩子进入地区天主教中学或是在公立学校上学,并不能被合理地想像成是一种随机的选择。尽管父母们会询问孩子的意见,但常常单独作决定;他们有关教育的决定是根据一些因素作出的,诸如他们预期的收入、他们倾向的宗教群体和意识形态、他们对自己孩子智力能力的估计、当地中学的相关声望等等。父母和孩子并非是惟一涉及的因素。教区中学对那些他们希望接纳的学生的成绩评判、学生挑选等行为都是以一些基本的因素为评判基础的,比如早期的教育经历,孩子或家长所希望的课程,以及孩子在行为上的声望。这些因素中的一部分可能会和中学的成绩有关系,在某种程度上,这些非独立的因素对成绩的影响将会把有关教区学校的学生成绩和公立学校的学生成绩的简单比较混淆起来。

很明显,走出这种困境的一个办法是用比较恒定的统计方法去衡量这些与成绩和学校选择相关的因素。然而,这样做存在的困难也是非常明显的。首先,必须明确这些相关因素。由于缺乏事实经验依据来支持,这项工作变得很困难。第二,假如选择的某个元素是这些因素(例子中提过到的)当中的一个,它就无法成为恒定的量,因为某种选择的发生是为某个群体而不是为其他群体。就目前情况而言,非天主教成员无法选择送他们的孩子去教会学校上学(为获得更加详细的有关 Coleman 在这方面的评论,请参阅:Rossi & Wright,1982)。

一个潜在的、富有成效的解决方法被计量经济学家提出(Goldberger 1980,Barnow et al. 1980,Berk &Ray 1982)。他们建议研究者建立一个清晰的有关决策过程的模型,同时把这种模型作为一个可以拥有衡量、用于自我选择的方法合并到结构平衡系统中去。尽管这种建议在某种程度上比通常用来把因变量加入回归方程的方法更具有吸引力,但仍然在很大程度上不切题,因为除了在一些特殊的情况下外,它很难建立确切的决策模型。

另一个用于非实验评价的非常重要的发展是时间序列模型在方法论上的运用,这个模型用来评价大规模项目的净相关[这种模型最开始是在经济预测中发展起来的(Pindyck & Rubinfeld 1976),接着特别地应用到评价问题中(McCleary & Hay 1980,Cook & Campbell 1979)]。它首先是被 Campbell &Stanley(1966)作为"断续时间序列"(interrupted time series)来设计,这种时间序列模型的应用使得评价新的大规模项目的效果或修改现成模型产生的影响成为可能,而不用求助于传统的随机实验。但这种方法还是受制于某些项目:这些项目关于它们可以获得的结果方面需要很长时间序列的数据,它们开始发生作用的时间必须及时确定。举个例子,比如新立法的实施。

那些最有名的"断续时间序列"评价多是有关马萨诸塞州巴特里—福克斯枪支法的评价(G. . L . Pierce &Bowers 1979,Deutsch &Alt 1977,Hay &McCleary 1979)。这部法律对没有许可证而持枪者采取强制惩罚,这样可以有效地减少枪案。使用时间序列模型,研究者制造模型来比较在这部法律未实施前的与枪支有关的案件发生走势和这部法律实施之后的枪案发生走势。研究结果表明这部法律对减少使用枪支犯罪只能

起到一点点作用。使用的时间序列模型(Box-Jenkins 模型)由相互连带和隶属的框架组成,在对随时间变化的运行程序的假设上,每一部分都是不同的。在某种程度上,不同模型的选择是一个判断问题,这就导致不同学者间有了关于法律真正效果的争论(比如:Hay & McCleary 1979, Deutsch 1979)。

刚才讨论过的两点进步暗含着社会学本质上不仅仅是用来评价研究。评价研究中自我选择问题的概念化已经很直接应用到许多社会学研究中了,这种研究依赖于纵向空间研究。由于遇到的数据分析问题是相同的,因此,在评价领域发展起来的解决方法,很快应用到各种因自我选择论题使得对结果的解释复杂化了的社会学研究中去。

在许多社会学家感兴趣的独立存在领域上,有关时间序列的关键数据是可以获得的。最大程度能得到的犯罪率数据可以追溯到约 50 年前;失业率方面几乎可以获得近 40 年的每月数据,等等。

对项目工具的研究是公共管理的重要内容。尽管好的例子非常少,但在原则上,去测试几种供选择的用于传递完成某种项目的方法并不比去测试几种选择性的项目困难。事实上,两个问题在形式上差不多。虽然项目实施被广泛认识到非常重要(Williams & Elmore 1976, Pressman & Wildavsky 1973, W. S. Pierce 1981),但是在相关项目中这些重要的研究并没有受到应有的重视。

注 释

1. 1964 年的初等和中等学校教育法案(McLaughlın 1975),1970 年出台的住房和城市发展预算要求对计划的房屋津贴项目进行评价(Struyk & Bendıck 1981),以及为劳工全面职业培训项目建立的可能性法案(Rossi et al. 1980)都出台了一些非常重要的评价。今天评价研究应用在社会干预的很多主要领域,包括健康、精神健康、犯罪审判、住房以及残疾儿童和他们的家庭。国防部使用评价研究的次数也越来越多。

2. 一张有关"评价研究会"成员的主要学科表格清晰地展现出评价研究领域的学科交叉性。关于成员学科领域的细目分类表显示:心理学 47%,社会学 10%,经济学 4%,政治学 6%,教育学 15%,其他 18%。

3. 由于时局困难,部分处于繁荣时期的壮观成就已经变的非常小了。在最高峰的时候,Abt 协会比任何一所波士顿的大学雇佣了更多社会科学方面的博士,也比许多联合学校要多。在过去的几年里,它的博士工作人员减少了大约 50%。

评价研究的未来

Delbert C. Miller

由于上面的关于评价研究的评价是由 Rossi 和 Wright 在 1984 年写的,评价研究已经经历了在财政支持上的下降。这种下降在 1980 年开始变得明显,是受里根统治的政治风暴影响。美国总审计局的委员会近来公布如下:

- 在 1984 至 1980 年间,所有的代理评价群体的专业人员数消减了 22%,从约1 500 人下降到约 1 200 人。相反,这段期间在这些机构的总人数只下降了 6%。
- 在 1984 至 1980 年间,用在项目评价上的资金消减了 37%,相比较之下,整个代理机构的资金却增加了 4%。
- 信息损失和结果歪曲总是被报道为是缺失评价的结果。这种失败在很多领域

非常严重，比如在防卫、环境以及劳工和职员等领域中。

不过，尽管近来形势不妙，但评价研究的未来仍充满希望：

> 减少州政府的财政赤字和鼓励公众对州政府的信心，是马上就要来临的国会和行政部门必须面临的两个最重要的问题。而对这两个问题具有决定意义的事是拥有获得及时的、技术性的正确信息能力，以便为立法、项目管理和公众意识服务。给第一听众——国会——的信息要回答关于资金如何运营以及能达到怎样成绩的问题；给第二听众——项目管理者——的信息要回答关于如何做才能符合法律以及达到运行的最高效果和最好效率；给第三听众——公众——的信息必须回答关于如何获得它们自身资金的问题。
>
> 对于这三个听众，项目评价是给它们提供消息的最重要工具(Chelimasky et al. 1989,p.25)。

这些需求在将来会得到满足。假如任何事物都在增加，比如传统的社会需求在急剧的增加，新的社会需求也在增加，预算数十亿地增加，那么评价研究就会变得对国会、对项目管理者以及对公众非常的重要。对评价的需求也就不仅仅限于州政府，它对于国家和城市政府同样重要。

评价已经成为政府工具的一部分。私人研究代理将会继续得到重要的项目测评合同。因此，对政府职员要有良好社会科学的训练以便承担研究项目的需求将会继续存在，社会学家继续可以在评价领域找到工作。

资料来源：Peter and James D. Wright(1984),"Evaluation Research Assessment,"*Annual Review of Sociology*,10,332 - 352，使用获得授权。

3.7.2 内外部评价的比较

按照Sonnichsen(2000)书中所说，评价是收集和分析证据的过程，接着把研究结果分发给确定的受众对象，这样政策和程序化的判断及决定就可以做出来了。

传统上，许多评价本质上是外部化的，是由这个组织之外选出的评价者做的，但是现在已经改变了。在今天预算紧缩的情况下，内部评价的一个最主要的优点就是低成本的支出。当然，一些其他的因素也在起着作用。

在下面，Sonnichsen(2000)给出了有关内部评价和外部评价的比较，包括优点和缺点。

表 3.1 内部评价与外部评价比较表

	内部评价	外部评价
优点	·为组织承担义务 ·对组织人事和运行程序的了解 ·对评价需求有快速的反应 ·作为组织成员具有更高的可信度 ·起到制度存储器的功能 ·经常与客户接触 ·支持决策制订的过程 ·容易获得数据 ·低成本 ·具有更强的能力去遵守组织的操作程序 ·面对临时的需求,具有可以承担除评价以外任务的弹性 ·有较强的能力去监督建议实施的过程 ·可以潜在地培养组织去关注价值评价 ·可以转换代理人角色进行服务 ·可以持续地评价	·可能拥有优秀的评价技巧 ·可能更具有独立性 ·给组织问题带来新的视角 ·更客观 ·较少受选举的影响 ·可以客观地评价组织各项目,包括作为参与人员的内部评价者
缺点	·可能在组织中缺乏影响力 ·可能缺乏自主性 ·陷入道德困境 ·会有额外工作负担 ·受到组织偏见的影响 ·可能缺乏技术评价上的专家	·缺乏对组织的了解 ·对组织数据的获得具有限制性 ·费用更高

参考文献

Sonnichsen, R. (2000). *High impact internal evaluation: A practitioner's guide to evaluating and consulting inside organizations.* Thousand Oaks, CA: Sage.

3.7.3 评价序列

无论评价者做了什么方面的评价(比如项目、政策、组织、产品,或者个体),并且无论是内部还是外部评价,评价的顺序基本是相同的。Sonnichsen(2000)列出了一系列的步骤(参阅上面的表格),可能它们中的大部分会在任何一个内部评价中出现,也同样会在外部评价中出现。

这种"内部评价必需的重要成分的方法上的序列"几乎保证了从头到尾的和完全的整套结果。以下的序列,根据个人手头的特殊问题可以很容易地调整,这是非常好的开始。

内部评价的重要成分

Richard Sonnichsen

1. 研究开始
   - 起始备忘录。包含充足的信息以便传递给感兴趣的资金拥有者和评价成员，这是评价的开始。
   - 权威人士
   - 目标
   - 期限
   - 组织成员
   - 管理者的要求
   - 电脑控制的日志记录
   - 约见有影响的经营主管人员和管理者
   - 关于部门头目和合适管理者的告示
   - 与项目专家联系
2. 文献收集
   - 回顾组织的政策和文件工作以及有信度的文献
3. 简洁陈述阶段
   - 展开论题,同时构架评价问题
4. 评价计划
   - 准备工作计划
   - 设计评价
   - 准备设计矩阵
5. 数据搜集和分析
   - 确定数据的质量和可获得性
   - 确定评价中数据需要达到的精确性
   - 在数据收集策略中解决所需数据的数量和质量
   - 决定是随机或是有目的的样本收集方法
   - 计划并调配好适当的统计和分析工具,以便得到预期的数据
6. 交流评价结果
   - 决定是否写书面报告
   - 准备合适的简本
   - 解决消化评价结果所需要的过程
7. 书写建议书
   - 确定选择是否恰当
8. 结束整个过程
   - 准备文案工作

- 保持和部署文案工作
- 解释注释
- 如果有必要,整理用过的材料
- 确保日志更新

9. 后续
   - 六个月后,确定根据建议产生的变化的状态以及被认可的建议,同时,试图开始测量评价的效果

资料来源:Richard Sonnichsen(2000),*High Impact Internerl Evaluation: A Practitioner' s Guide to Evaluating and Consulting Inside Orgnizations* (Thousand, CA: Sage),使用获得了授权。

3.7.4 在评价过程中图表的概念

John Owen 和 Patricia Rogers(1999)已经创造出他们所提及的作为评价的"元模型"(meta-model),这个模型建立在五个评价图表的基础上,每个都是"在一组普通的问题中有确定的方向和重点,它们为调查的计划和实施提供了指导。"

这五个种类或图表是:

- (心理)前摄(proactive)
- 澄清(clarificative)
- 互动(interactive)
- 监控(monitoring)
- 影响(impact)

正如你将要看到的,它们能够通过一系列的标准进行比较,这些标准包括评价的目标,不同图表表达的问题,评价中使用的方法,评价中的重点,以及其他。表格3.2是这五种评价图表组成的,它们反映了这些标准的作用。

参考文献

Owen, John M. , & Rogers, Patricia. (1999). *Program evaluation: Forms and approaches.* London: Sage.

表 3.2 评价表格：定位、典型问题以及关键方法

	前摄评价	澄清评价	互动评价	监督评价	效果评价
定位	综合	澄清	改进	辩护/归位	辩护/责任
典型问题	· 做这个项目是否有必要？ · 对这个项目将要做的事情，我们了解多少？ · 在这个领域最好的实践是什么？ · 是否已经尽力去寻找在这个问题上可能存在的其他方法？ · 与这个问题相关的研究或传统谚语可以告诉我们什么？ · 从外部信息到更新已经存在的政策或项目，我们可以发现什么？	· 期望得到什么样的结果，如何去实现它们？ · 这个项目隐藏的原则是什么？ · 项目的哪些成分需要修改，以便最大化期望的结果？ · 项目是否似是而非？ · 这个项目的哪些方面适用于一系列监控或效果评价？	· 这个项目试图达到什么结果？ · 这个服务是如何进行的？ · 项目交付是否在进行？ · 交付是否和项目计划同步进行？ · 如何做可以使组织更有效？	· 项目是否达到目标总体？ · 实施是否符合项目基准？ · 项目在不同点之间是如何进行的？ · 与一个月前相比实施进行得如何？ · 项目成本是在上升还是下降？ · 我们可以如何调整项目使得它更有效？ · 我们可以如何调整项目使得它更易推动？ · 为了保证更有效的陈述，是否存在需要关注的项目点？	· 项目是否按照计划进行？ · 计划要求的目标是否已经达到？ · 项目服务的需要是否已经达到？ · 臆想不到的结果是什么？ · 执行策略是否得到设想的结果？ · 执行中的差异是如何影响项目结果的？ · 项目是否已经有很大的成本？

续表

	前摄评价	澄清评价	互动评价	监督评价	效果评价
关键方法	·需求评价 ·研究回顾 ·最佳实践的回顾(标杆,benchmarking)	·价值评价 ·逻辑/理论发展 ·信誉	·反应力 ·行动研究 ·定性评论 ·发展性 ·权力/效力	·成分分析 ·转移成果评价 ·系统分析	·客观基础 ·过程—结果分析 ·基础需求 ·目标自主 ·绩效审核
项目状态	无	发展	发展	成形的	成形的
中心关注点	项目的前后关系	所有的成分	传递	传递/结果	传递/结果
时间选择(vis-à-vis 项目传递)	以前	期间	期间	期间	期后
证据综合	回顾文献和数据,参观地点,以及其他一些互动的方法。中心群体、名义群体,以及对测评需要非常有用的特尔菲方法	通常依赖文献分析、回顾和观察的整合。发现的结果包括项目计划和组织的实现。 可以改善民意	依赖清晰的实地研究,包括观察。数据结构程度取决于方法。可以求助于提供者和项目参与者	系统方法要求处理信息系统(MIS)具有可获得性,指标的使用,成果信息的有意义使用	在可能使用数据处理和控制组,以及使用测量和定量数据的情况下,传统上要求使用预定的研究设计。要完成研究一般都需要观测性数据。决定所有结果要求使用更多探索性的方法和运用定性的证据

3.7.5 关于评价研究的推荐读物

Sage 出版公司长期以来致力于评价研究图书的出版,其出版物的范围涉及评价的理论、方法及应用等。对于评价的操作方面感兴趣的学生及研究者应首选阅读“项目评价工具包”系列作品,并进一步阅读这一部分所罗列的其他一些针对具体问题的评价的一些推荐读物。这里提供的这个全面的参考资料,是为那些想较全面地理解这一问题的学者提供指导的。

(1)项目评价工具包(Sage Publications)

“项目评估工具包”(*Program Evaluation Kit*)最初出版于 1978 年,由 Joan L. Herman 任主编。其中为不同类型项目的评价提供了多种不同的技术。

- Volume 1—*Evaluator's Handbook* (1988) by Joan L. Herman, Lynn Lyons Morris, and Carol Taylor Fitz-Gibbon. This first volume is the heart of the Program Evaluation Kit and provides a broad overview of evaluation planning and a practical guide to designing and managing programs.
- Volume 2—*How to Focus an Evaluation* (1988) by Brian Stecher and W. Alan Davis. This volume provides a broad overview of evaluation planning and a practical guide to designing and managing programs.
- Volume 3—*How to Design a Program Evaluation* (1988) by Carol Taylor Fitz-Gibbon and Lynn Lyons Morris. This volume reflects the tremendous explosion of interest in this vital area of the evaluation process and recognizes that deciding what to evaluate is a complex negotiation process that involves many different factors.
- Volume 4—*How to Use Qualitative Methods in Evaluation* (1988) by Michael Quinn Patton. Introduces the reader to qualitative approaches.
- Volume 5—*How to Assess Program Implementation* (1988) by Jean A. King, Lynn Lyons Morris, and Carol Taylor Fitz-Gibbon. Extensively revised to reflect modern views of program implementation, this volume introduces the variety of functions served by implementation studies and the roles played by qualitative and quantitative data.
- Volume 6—*How to Measure Attitudes* (1988) by Marlene E. Henerson, Lynn Lyons Morris, and Carol Taylor Fitz-Gibbon. An important part of any evaluation process, this book focuses on the assessment of attitudes.
- Volume 7—*How to Measure Performance and Use Tests* (1988) by Lynn Lyons Morris, Carol Taylor Fitz-Gibbon, and Elaine Lindheim. The evaluator's role in performance measurement is a critical element, and this volume focuses on ways an evaluator can select, develop, and analyze tests.
- Volume 8—*How to Analyze Data* (1988) by Carol Taylor Fitz-Gibbon and Lynn Lyons Morris. This is a basic introduction to a variety of elementary statistical techniques, including those for summarizing data, for examining differences between

groups, and for examining relationships between two measures.

- Volume 9—*How to Communicate Evaluation Findings* (1988) by Lynn Lyons Morris, Carol Taylor Fitz-Gibbon, and Marie E. Freeman. This volume includes examples from a wide range of disciplines and shows the reader how to communicate results to users and stakeholders throughout the evaluation process.

(2) Sage 公司在评价方面的其他出版物

《评价研究年鉴》(*The Evaluation Studies Review Annuals*)由 Sage Publications 出版。相关信息可以在以下网站上找到:www. sagepub. com.

- Volume 1: edited by Gene V. Glass (1976, 704 pages)
- Volume 2: edited by Marcia Guttentag with Shalom Saar (1977, 736 pages)
- Volume 3: edited by Thomas D. Cook and Associates (1978, 783 pages)
- Volume 4: edited by Lee Sechrest, Stephen G. West, Melinda A. Phillips, Robin Redner, and William Yeaton (1979, 768 pages)
- Volume 5: edited by Ernst W. Stromsdorfer and George Farkas (1980, 800 pages)
- Volume 6: edited by Howard E. Freeman and Marian A. Solomon (1981, 769 pages)
- Volume 7: edited by Ernest R. House and Associates (1982, 752 pages)
- Volume 8: edited by Richard J. Light (1983, 672 pages)
- Volume 9: edited by Ross F. Conner, David G. Altman, and Christine Jackson (1984, 752 pages)
- Volume 10: edited by Linda H. Aiken and Barbara H. Kehrer (1985, 650 pages)
- Volume 11: edited by David S. Cordray and Mark W. Lipsey (1986-1987, 757 pages)
- Volume 12: edited by William R. Shadish, Jr., and Charles S. Reichardt (1988, 704 pages)

(3)评价研究的精选实例

Covello, Vincent. (Ed.). (1980). *Poverty and public policy: An evaluation of social science research.* Rochester, NY: Schenkman.

Creemers, B. P. (Ed.). (1985). *Evaluation research in education.* Philadelphia, PA: Taylor and Francis.

Fults, Gail J. (1992). *How participatory evaluation research affects management control process of a multinational nonprofit organization.* New York: Garland.

Glaser, Daniel. (1987). *Evaluation research and decision guidance for correctional, addiction treatment, mental health, education, and other people changing agencies.* Piscataway: Transaction Publishers.

Houser, Rick. (1998). *Counseling and educational research: Evaluation and application.* Thousand Oaks, CA: Sage.

Irden, Ellen R. (1996). *Evaluation research articles from start to finish.* Thousand Oaks,

CA: Sage.

Lomand, Turner C. (1999). *Social science research: A cross section of journal articles for discussion and evaluation.* Los Angeles: Pyrczak Publishing.

Malaney, Gary D. (Ed.). (1999). *Student affairs research, evaluation, and assessment: Structure and practice in an era of change.* San Francisco: Jossey-Bass.

Pollard, William. (1986). *Bayesian statistics for evaluation research: An introduction.* Beverly Hills, CA: Sage.

Swerver, S. (1995). *Climate change research: Evaluation and policy implications.* New York: Elsevier Science.

Teichler, Ulrich, & Maiworm, Friedhelm, (1997). *The ERAMUS experience: Major findings of the ERAMUS Evaluation Research Project.* Brussels, Belgium: Commission of the European Communities.

Vance, Mary. (1987). *Evaluation research (social action programs): Monographs.* Monticello, VA: Vance Bibliographies.

(4)评价研究总体参考文献

Boggio, Giorgio (Ed.). (1982). *Evaluation of research and development.* Norwood, NJ: Kluwer Academic.

Buchanan, Nina K., & Feldh usen, John E (Eds.). (1991). *Conducting research and evaluation in gifted education: A handbook of methods and applications.* New York: Teachers College Press.

Burgess, Robert G. (1992). *Educational research and evaluation: For policy and practice?* Philadelphia, PA: Taylor and Francis.

DIANE Publishing Company. (1992). *Drugs in the workplace: Research and evaluation data.* Collingdale: Author.

Fink, Arlene. (1992). *Evaluation fundamentals: Improving health programs, research, and policies.* Thousand Oaks, CA: Sage.

Frizzell, Leon A. (1984). *Evaluation of research reports: Bioeffects literature reviews.* Laurel: American Institute of Ultrasound in Medicine.

Hadley, Robert G., & Mitchell, Lynda K. (1994). *Counseling research and program evaluation.* Pacific Grove, CA: Brooks/Cole.

Hoorweg, Jan, & Neimeijer, Rudo. (1988). *Intervention in child nutrition: Evaluation studies in Kenya.* New York: Routledge.

Ladbury, John E. (1999). *Evaluation of the NASA Langley Research Center Mode-Stirred Chamber Facility.* Washington, DC: Government Printing Office.

Meyer, Michael M., & Fienberg, Stephen E. (Eds.). (1992). *Assessing evaluation studies: The case of bilingual education.* Ann Arbor, MI: Books on Demand.

Milinki, Andrea K. (Ed.). (1999). *Cases in qualitative research: Research reports for discussion and evaluation.* Los Angeles: Pyrczak Publishing.

Oman, Ray C., & Masters, Bill. (1987). *Implementing change in organizations, based on analysis and evaluation studies: A bibliography.* Monticello, VA: Vance Bibliographies.

Pond, D. J. (1998). *Evaluation of human factors research for ultrasonic inservice inspection.* Washington, DC: Government Printing Office.

Rudman, Jack. (1994). *Director of research and evaluation.* Sysosset: National Learning Corporation.

Schulenburg, J. Matthias. (Ed.). (in press). *The influence of economic evaluation studies on health care decision making: A European survey.* Burke, CA: IOS Press.

Wells, Kathleen. (1991). *Family preservation services: Research and evaluation.* Thousand Oaks, CA: Sage.

Windsor, M., Lange, K., & Mohr, V. (1994). *Evaluation of the Fisheries and Aquaculture Research Programme*, 1988-1992. Brussels: Commission of the European Communities.

(5)评价杂志

以下是由 Sage Publications 出版的评价方面的杂志。具体信息可以在以下网站中看到:www. sagepub. com.

- *Evaluation: The International Journal of Theory, Research, and Practice*
- *Evaluation and the Health Professions*
- *Evaluation Review: A Journal of Applied Social Research*

下面是由其他有关机构出版的评价方面的杂志:

- *Evaluation Practice* (JAI Press)
- *New Directions for Evaluation* (Jossey-Bass)
- *Evaluation and Program Planning* (Pergamon)
- *Journal of Policy Analysis and Management* (John Wiley)
- *Canadian Journal of Program Evaluation* (University of Calgary Press)
- *Evaluation Journal of Australia* (Australian Evaluation Society)
- *Educational Evaluation and Policy Analysis* (American Educational Research Association)
- *Assessment and Evaluation in Higher Education* (Carfax Publishing Ltd.)

(6)调查工具包

- Bourque, L., & Fielder, E. P. *How to conduct self-administered and mail surveys.*
- Fink A. *How to analyze survey data.*
- Fink A. *How to ask survey questions.*
- Fink A. *How to design surveys.*
- Fink A. *How to report on surveys.*
- Fink A. *How to sample in surveys.*
- Fink A. *The survey handbook.*

- Fink A. *The survey kit.*
- Frey, J. , & Oishi, S. M. *How to conduct interviews by telephone and in person.*
- Litwin, M. *How to measure survey reliability and validity.*

(7)评价者及政策分析者的专业组织

- American Evaluation Association at www. eval. org
- Association for Public Policy Analysis and Management at www. appam. org
- Canadian Evaluation Association at www. evaluationcanada. ca/
- Australian Evaluation Association at www. aes. asn. au
- European Evaluation Association at www. europeanevaluation. org
- United Kingdom Evaluation Association at www. evaluation. org. uk
- Italian Evaluation Association at www. valutazione. it

3.7.6 评价方面的百科全书

Airasian, Peter W. , & Gullickson, Arlen R. (1997). *Teacher self-evaluation tool kit.* Thousand Oaks, CA: Corwin Press.

Awasthi, Dinesh N. , & Sebastian, Jose. (1996). *Evaluation of entrepreneurship development programmes.* New Delhi: Sage India.

Barker, Cornelius L. , & Searchwell, Claudette J. (1998). *Writing meaningful teacher evaluations—right now! The principal's quick start reference guide.* Thousand Oaks, CA: Corwin Press.

Berk, Richard A. , & Rossi, Peter H. (1990). *Thinking about program evaluation.* Newbury Park, CA: Sage.

Bickman, Leonard, & Rog, Debra J. (1995). *Children's mental health services: Research, policy, and evaluation.* Thousand Oaks, CA: Sage.

Blenkin, Geva M. , & Kelly, A. V. (1994). *The National Curriculum and early learning: An evaluation.* London: Paul Chapman Publishing.

Borders, L. Dianne, & Drury, Sandra M. (1992). *Counseling programs.* Thousand Oaks, CA: Corwin Press.

Boruch, Robert F. (1997). *Randomized experiments for planning and evaluation: A practical guide.* Thousand Oaks, CA: Sage.

Bright, George W. , Uprichard, A. Edward, & Jetton, Janice H. (1993). *Mathematics programs: A guide to evaluation.* Thousand Oaks, CA: Corwin Press.

Card, Josefina J. (1993). *Handbook of adolescent sexuality and pregnancy: Research and evaluation instruments.* Newbury Park, CA: Sage.

Chattopadhyay, Manabendu, Maiti, Pradip, & Rakshit, Mihir. (1996). *Planning and economic policy in India: Evaluation and lessons for the future.* New Delhi: Sage India.

Chelimsky, Eleanor, & Shadish, William R. (1997). *Evaluation for the 21st century: A handbook.* Thousand Oaks, CA: Sage.

Chen, Huey-Tsyh. (1994). *Theory-driven evaluations.* Thousand Oaks, CA: Sage.

Clarke, Alan. (1999). *Evaluation research: An introduction to principles, methods and practice.* London: Sage.

Dale, Reider. (1998). *Evaluation frameworks for development programmes and projects.* New Delhi: Sage India.

Edwards, Ward, & Newman, J. Robert. (1982). *Multiattribute evaluation.* Beverly Hills, CA: Sage.

Evaluation and the health professions [journal published by Sage Publications].

Evaluation review: A journal of applied social research [journal published by Sage Publications].

Evaluation: The international journal of theory, research, and practice [journal published by Sage Publications].

Fetterman, David M. (2000). *Foundation of empowerment evaluation.* Thousand Oaks, CA: Sage.

Fetterman, David M., Kaftarian, Shakeh J., & Wandersman, Abraham. (1995). *Empowerment evaluation: Knowledge and tools for self-assessment and accountability.* Thousand Oaks, CA: Sage.

Fink, Arlene. (1993). *Evaluation fundamentals: Guiding health programs, research, and policy.* Thousand Oaks, CA: Sage.

Fink, Arlene. (1995). *Evaluation for education and psychology.* Thousand Oaks, CA: Sage.

Finkel, Martin A., & Giardino, Angelo P. (Eds.). (2001). *Medical evaluation of child sexual abuse: A practical guide* (2nd ed.). Thousand Oaks, CA: Sage.

Fitz-Gibbon, Carol Taylor, & Morris, Lynn Lyons. (1988). *How to design a program evaluation.* Newbury Park, CA: Sage.

Fowler, Floyd J., Jr. (1995). *Improving survey questions: Design and evaluation.* Thousand Oaks, CA: Sage.

Freeman, Howard E., Rossi, Peter H., & Sandefur, Gary D. (1993). *Workbook for evaluation: A systematic approach.* Newbury Park, CA: Sage.

Giardino, Angelo P., Christian, Cindy W., & Giardino, Eileen R. (1997). *A practical guide to the evaluation of child physical abuse and neglect.* Thousand Oaks, CA: Sage.

Giardino, Angelo P., Finkel, Martin A., Giardino, Eileen R., Seidl, Toni, & Ludwig, Stephen. (1992). *A practical guide to the evaluation of sexual abuse in the prepubertal child.* Newbury Park, CA: Sage.

Gil, Libia. (2001). *Principal peer evaluation: Promoting success from within.* Thousand Oaks, CA: Corwin Press.

Glasman, Naftaly S. (1994). *Making better decisions about school problems: How administrators use evaluation to find solutions.* Thousand Oaks, CA: Corwin Press.

Gould, Jonathan W. (1998). *Conducting scientifically crafted child custody evaluations.* Thousand Oaks, CA: Sage.

Grembowski, David. (2001). *The practice of health program evaluation.* Thousand Oaks, CA: Sage.

Guba, Egon G., & Lincoln, Yvonna S. (1989). *Fourth generation evaluation.* Thousand Oaks, CA: Sage.

Hambleton, Robin, & Thomas, Huw. (1995). *Urban policy evaluation: Challenge and change.* London: Paul Chapman Publishing.

Herman, Joan L. (1988). *Program evaluation kit.* Newbury Park, CA: Sage.

Herman, Joan L., & Winters, Lynn. (1992). *Tracking your school's success: Aguide to sensible evaluation.* Thousand Oaks, CA: Corwin Press.

Hilton, N. Zoe. (1993). *Legal responses to wife assault: Current trends and evaluation.* Newbury Park, CA: Sage.

Holloway, Jacky, Lewis, Jenny, & Mallory, Geoff. (1995). *Performance measurement and evaluation.* London: Sage.

House, Ernest R. (1993). *Professional evaluation: Social impact and political consequences.* Newbury Park, CA: Sage.

House, Ernest R., & Howe, Kenneth R. (1999). *Values in evaluation and social research.* Thousand Oaks, CA: Sage.

Houser, Rick. (1998). *Counseling and educational research: Evaluation and application.* Thousand Oaks, CA: Sage.

Imrie, Rob, & Thomas, Huw. (1999). *British urban policy: An evaluation of the urban development corporations.* London: Sage.

Jenny, Carole. (1996). *Medical evaluation of physically and sexually abused children.* Thousand Oaks, CA: Sage.

Kosecoff, Jacqueline, & Fink, Arlene. (1983). *Evaluation basics: A practitioner's manual.* Beverly Hills, CA: Sage.

Kushner, Saville. (2000). *Personalizing evaluation.* London: Sage.

Love, Arnold J. (1991). *Internal evaluation: Building organizations from within.* Newbury Park, CA: Sage.

Millman, Jason. (1981). *Handbook of teacher evaluation.* Beverly Hills, CA: Sage.

Millman, Jason. (1997). *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* Thousand Oaks, CA: Corwin Press.

Millman, Jason, & Darling-Hammond, Linda. (1991). *The new handbook of teacher evaluation: Assessing elementary and secondary school teachers.* Thousand Oaks, CA: Corwin Press.

Mohr, Lawrence B. (1996). *Impact analysis for program evaluation.* Thousand Oaks, CA: Sage.

Morris, Lynn Lyons, Fitz-Gibbon, Carol Taylor, & Freeman, Marie E. (1988). *How to communicate evaluation findings.* Newhury Park, CA: Sage.

Mullins, Terry W. (1994). *Staff development programs: A guide to evaluation.* Thousand Oaks, CA: Corwin Press.

Mundry, Susan, Britton, Edward, Raizen, Senta, & Loucks-Horsley, Susan. (2000). *Designing successful professional meetings and conferences in education: Planning, implementation, and evaluation.* Thousand Oaks, CA: Corwin Press.

Newman, Dianna L., & Brown, Robert D. (1996). *Applied ethics for program evaluation.* Thousand Oaks, CA: Sage.

Olson, Mary W., & Miller, Samuel D. (1993). *Reading and language arts programs: A guide to evaluation.* Thousand Oaks, CA: Corwin Press.

O'Sullivan, Rita G., & Tennant, Cheryl V. (1993). *Programs for at-risk students: A guide to evaluation.* Thousand Oaks, CA: Corwin Press.

Owen, John M., & Rogers, Patricia. (1999). *Program evaluation: Forms and approaches.* London: Sage.

Padaki, Vijay. (1995). *Development intervention and programme evaluation: Concepts and cases.* New Delhi: Sage India.

Patton, Michael Quinn. (1983). *Practical evaluation.* Beverly Hills, CA: Sage.

Patton, Michael Quinn. (1988). *How to use qualitative methods in evaluation.* Newbury Park, CA: Sage.

Patton, Michael Quinn. (1991). *Family sexual abuse: Frontline research and evaluation.* Newbury Park, CA: Sage.

Patton, Michael Quinn. (1997). *Utilization-focused evaluation: The new century text.* Thousand Oaks, CA: Sage.

Patton, Michael Quinn. (2002). *Qualitative research and evaluation methods.* Thousand Oaks, CA: Sage.

Pawson, Ray, & Tilley, Nicholas. (1997). *Realistic evaluation.* London: Sage.

Peterson, Kenneth D. (2000). *Teacher evaluation: A comprehensive guide to new directions and practices* (2nd ed.). Thousand Oaks, CA: Corwin Press.

Pietrzak, Jeanne, Ramler, Malia, Renner, Tany, Ford, Lucy, & Gilbert, Neil. (1990). *Practical program evaluation: Examples from child abuse prevention.* Newbury Park, CA: Sage.

Robson, Colin. (1999). *Small scale evaluations: Principles and practice.* London: Sage.

Rossi, Peter H., Freeman, Howard E., & Lipsey, Mark W. (1999). *Evaluation* (6th ed.). Thousand Oaks, CA: Sage.

Rutman, Leonard, & Mowbray, George. (1983). *Understanding programme evaluation.* London: Sage.

Sanders, James R., & The Joint Committee on Standards for Educational Evaluation.

(1994). *The Program Evaluation Standards: How to assess evaluations of educational programs.* Thousand Oaks, CA: Sage.

Schinke, Steven P., Botvin, Gilbert J., & Orlandi, Marlo A. (1991). *Substance abuse in children and adolescents: Evaluation and intervention.* Newbury Park, CA: Sage.

Schram, Barbara. (1997). *Creating small scale social programs.* Thousand Oaks, CA: Sage.

Scriven, Michael. (1991). *Evaluation thesaurus.* Newbury Park, CA: Sage.

Shadish, William R., Cook, Thomas D., & Leviton, Laura C. (1993). *Foundations of program evaluation: Theories of practice.* Newbury Park, CA: Sage.

Shah, A. M., Baviskar, B. S., & Ramaswamy, E. A. (1996). *Social structure and change: Vol. 1. Theory and method—an evaluation of the work of M. N. Srinivas.* New Delhi: Sage India.

Shaw, Ian. (1999). *Qualitative evaluation.* London: Sage.

Shaw, Ian, & Lishman, Joyce. (1999). *Evaluation and social work practice.* London: Sage.

Sonnichsen, Richard. (2000). *High impact internal evaluation: A practitioner's guide to evaluating and consulting inside organizations.* Thousand Oaks, CA: Sage.

Stahl, Philip Michael. (1994). *Conducting child custody evaluations: A comprehensive guide.* Thousand Oaks, CA: Sage.

Stahl, Philip Michael. (1999). *Complex issues in child custody evaluation.* Thousand Oaks, CA: Sage.

Stecher, Brian, & Davis, W. Alan. (1988). *How to focus an evaluation.* Newbury Park, CA: Sage.

Stufflebeam, Daniel L. (1988). *The personnel evaluation standards: How to assess systems for evaluating educators.* Thousand Oaks, CA: Corwin Press.

Torres, Rosalie T., Preskill, Hallie, & Piontek, Mary E. (1996). *Evaluation strategies for communicating and reporting: Enhancing learning in organizations.* Thousand Oaks, CA: Sage.

Vallecorsa, Ada L., deBettencourt, Laurie U., & Garriss, Elizabeth. (1992). *Special education programs: A guide to evaluation.* Thousand Oaks, CA: Corwin Press.

Veale, James. (2001). *Practical evaluation for collaborative services: Goals, processes, tools, and reporting systems for school-based programs.* Thousand Oaks, CA: Corwin Press.

Yuan, Ying-Ying T., & Rivest, Michele. (1990). *Preserving families: Evaluation resources for practitioners and policymakers.* Newbury Park, CA: Sage.

3.8 研究中的伦理及其实践

3.8.1 美国社会学会的伦理准则

研究伦理是,而且必须是从事研究的任何人最优先关心的内容。几乎所有的专业组织都有相应的从事研究的伦理原则。而且在任何专业训练的早期,这个伦理原则必须充分讨论,从而能使其任何组成部分都能为从事本专业研究的任何人所充分理解。下面就是美国社会学学会(ASA)的伦理准则。

美国社会学学会的伦理准则

绪　论

美国社会学学会的伦理准则设立了原则和道德标准,以此规范社会学家的专业责任和行为。这些原则和道德标准被视为衡量和检验每天发生的职业行为的准则。它们为社会学家的标准论述提供了依据,并为社会学家解决在专业工作中可能遇到困难提供了指导性建议。

美国社会学学会的伦理准则由绪论、导言、五个总原则和特殊伦理准则等组成。它也是美国社会学学会伦理委员会提出、调查和解决缺乏职业道德行为时的规则和程序的有机组成部分。

导言和总原则是充满激情的目标,引领着社会学家朝社会学最高理念而努力。尽管它们不是强迫性的规定,但它们应该被社会学家当作形成伦理行为的必须步骤,也可以被他们当作在解释伦理标准过程中的道德规范主体。

伦理准则阐明了社会学家可实施的行为准则。绝大多数伦理准则涉及面宽泛,以利于社会学家适应各种角色,而且任何一种伦理准则根据事情的来龙去脉,在实际中发生相应的变化。伦理准则不是详尽无遗的,没有被伦理准则特别提及的行为并不必然是道德或不道德的。

美国社会学学会的会员资格要求其成员遵守该学会的伦理准则和美国社会学学会道德委员会的政策和程序。在加入该协会时,成员会被告知必须遵守此要求,如果违反,将会被给予相应的处分,包括会员资格的终结。只有在这些伦理准则成为行为的一部分,或者它们影响着与工作相关的功能,或者行为本质上是社会学的,受此伦理准则约束的成员才可以被进行相关的审查。与社会学家职业角色上的行为无关或没有因果关系的个人行为,不受伦理准则的约束。

导 言

这个伦理准则把建立于社会学家专业而科学的工作之上的一套通用的价值观连接起来。它既要提供总原则，又要提供解决社会学家遇到种种职业情况的方法准则。作为首要目标，它为那些与社会学家一直工作的个人或群体提供福利和保护。它是每一位社会学家的责任，以此激发他们追求可能的最高标准来指导在研究、教学、实践与服务中的行为。

总原则

下列总原则对社会学家既是工作志向，又是处理各种情况下决定道德方向的指针。它们是职业行为最高理想的成例。

原则 A：专业竞争

社会学家在工作中努力获得最高的竞争水准，他们认识到自己经历中的局限，他们只从事那些与他们所受的教育、训练或经历相匹配的任务。他们认识到继续教育对他们保持专业竞争优势的必要，并且在专业行为中，恰当地利用所需科学的、专业的、技术的和行政的资源来确保竞争力。当要考虑到他们的学生、研究参与者和委托人的利益时，他们会与其他专家协商处理。

原则 B：诚实正直

社会学家是诚实和公正的，而且对于与他们一起从事专业行为——研究、教学、实践和服务——的其他人表示尊敬。他们不会故意地危害自己或其他专家的利益。他们处理自己的事情时要激发信任与自信，他们不会故意地发表错误、误导或欺骗的言论。

原则 C：专业而科学的责任

社会学家们坚守最高的科学而专业的标准，承担工作责任。他们深知，他们形成了一个共同体，应对其他社会学家表示尊重，即使是在他们不同意他人的理论、方法以及达到专业行为的个人途径的情况，也是如此。他们重视公众对社会学的信任，关心自己和其他社会学家的伦理行为是否会危及这种信任。尽管社会学家的努力通常是学院式的，但他们绝不能让他们学院式的愿望超越他们伦理行为而带来的责任。在适当的时候，他们与同事协商以达到防止或避免不道德行为的目的。

原则 D：尊重他人的权利、尊严和差异

社会学家尊重权利、尊严和尊敬他人。在专业行为中，他们努力减少偏见，不会容忍任何形式的基于年龄、性别、民族、种族、国别、地区、性别取向、残疾、健康状况，或者婚姻、家庭，以及父母亲状况的歧视。他们在服务、教育和研究各种有着典型特征的人群过程中，对文化、个体和角色差异非常敏感。在所有与工作有关的行为中，社会学家认识到他人拥有与自己不同的价值观、态度和观点的权利。

原则 E：社会责任

社会学家充分意识到他们对所居住和工作的社区和社会的专业而科学的责任。他们运用并且将自己的知识公之于众，从而造福于公众利益。当他们从事研究时，他

们努力提高社会学的科学程度并服务于公众利益。

伦理标准

1.专业及科学的标准

社会学家在研究、教学、实践和服务活动中，坚持运用那些最具合理及负责任的可能性的技术标准。他们科学而专业地依赖已有知识，诚实而正直地做事。在从事与工作有关的仪式或行动时，他们避免发表不正确、欺骗或没有根据的声明。

2.竞　争

(a)社会学家只是在他们能力范围内从事研究、教学、实践以及提供服务。这种能力由他们所受的教育、训练、管理经验，或者合适的专业经历所决定。

(b)社会学家只有在采取了合理步骤以确保在新领域的竞争后，才在这些领域或运用新技术来从事研究、教学、实践以及提供服务。

(c)从事研究、教学、实践和服务的社会学家通晓自己从事领域目前的科学及专业的信息。他们不断努力获得技能上的竞争力。

(d)当个人情况可能妨碍自己的专业工作，或者可能导致危害学生、被指导者、实验对象、客户、同事，或者那些拥有科学的、教育的、咨询的或者其他专业责任的人，社会学家就会避免从事这种行为。

3.陈述和误用专家意见

(a)在研究、教学、实践、服务，以及作出专家判断或表达专家意见的其他情形时，社会学家准确而公正地描述专家意见所涉及的领域和深度。

(b)社会学家不接受那些来自个人客户、组织客户或主办者的，看上去很有可能违反伦理原则中一些准则的准许、合同、咨询，或者工作委派。当社会学家发现有违反现象并且不能改正时，他们就应从这种行为中脱身出来。

(c)因为社会学家科学而专业的判断有可能影响他人的生活，所以他们时刻警醒而且提防个人的、财金的、社会的、组织的，或者政治的因素可能导致他们的知识、专家意见以及影响的误用。

(d)如果社会学家知道他们在工作中出现了误用或者错误的陈述，他们就会采取合理措施来改正或者尽量减少由此造成的损失。

4.授权和监督

(a)社会学家给他们的学生、被指导者，或者雇员提供正确的培训和监督，并采取合理步骤来确保这些人负责任地、合适地、符合伦理地提供服务。

(b)社会学家根据他们的学生、被指导者，或者雇员的知识、训练或者经验，只将那部分能合理预期被这些人独立或者在监督的范围内切实履行的职责，授权给他们。

5.不歧视

社会学家在工作中，不因年龄、性别、民族、种族、国别、地区、性别取向、残疾、健康状况、婚姻、家庭，或者父母亲状况，以及任何其他被法律禁止但可应用的偏见而产生歧视。

6. 保　密

(a) 无论是出于个人的、经济的,或者专业进步的理由,社会学家也不能将学生、被指导者、雇员,或者研究参与者暴露给对他们有着直接或间接的监督、评价或者其他权力的人。

(b) 社会学家不直接对任何有着性关系的人,包括学生、被指导者、雇员,或者研究参与者进行指导或者行使评价权力。

7. 骚　扰

社会学家不骚扰包括学生、被指导者、雇员或研究参与者在内的所有人。骚扰由一种单一的强烈而严厉的行动或者多种持续稳固或普遍深入的卑下的、辱骂的、无礼的,甚至制造一种敌对的专业人员或工作环境的行动组成。性骚扰可能包括性诱惑、身体接触、本质上是性的口头或非口头行为。因为个人的种族或种族划分而引起的种族骚扰可能包括不必要的、夸大的,以及无论是口头或非口头的无根据的关注或攻击。

8. 就业决定

在参与做与就业相关的决定时,在寻求就业机会或计划从某岗位辞职时,一个社会学工作者应当坚守高尚的伦理准则。

8.01　公平就业实践

(a) 在参与做与就业相关的决定时,社会学家应当努力确保所有的全职和兼职雇员都有公平的机会并得到同等对待。他们在雇佣、薪水、提升、待遇,或任何其他的就业条件方面,或以年龄、性别、民族、种族、国别、地区、性别取向、残疾、健康状况、婚姻、家庭,或者父母亲状况,或任何其他公然被排斥在法律保护之外的东西,在这些方面,他们不应被区别对待。

(b) 在参与做与就业相关决定时,社会学家详述雇佣、提升、使用期和终止雇佣的要求,而且与这些全职和兼职雇员及潜在的雇员就这些要求进行交流。

(c) 在参与做与就业相关决定时,社会学家有责任了解公平就业章程,并与雇员交流这些信息,而且为全职和兼职雇员营造一种赞成公平就业训练的氛围。

(d) 在参与做与公平就业相关决定时,社会学家应当告知可能的全职或兼职雇员研究和出版方面的限制,并且就可能限制研究和学术活动的条件达成清楚的理解。

8.02　雇员的责任

(a) 在寻求就业时,社会学家应当为雇主提供关于专业要求和经验的精确和完整的信息。

(b) 离开一个岗位时,无论是永久的还是暂时的,社会学家应当为雇主提供充分的信息,并且采取合理措施减少离岗的负面影响。

9. 利益冲突

社会学家在职业活动中应坚持高自尊水平,并避免利益冲突和卷入冲突。当社会学家的个人或经济利益使他们不能毫无偏见地从事专业工作时,冲突就会发生。在研究、教学、实践和服务中,社会学家对可以引起利益冲突的环境应非常警觉,他们应采取适当的行动避免冲突,或者对适当的各方揭示这个情况。

9.01 职业标准的坚持

要做到不受社会学家个人或经济利益，或他们的雇主以及委托人的利益的影响，社会学家应当在以下方面坚持专业和科学标准：

(a)资料收集、理论分析或解释数据。

(b)研究报告。

(c)教学、职业行为和社会学知识的大众传播。

(d)合适的契约、咨询、服务活动的确认和完成。

9.02 公开的事物

社会学家应当公开财政支持的相关来源，以及相关的个人或职业关系，这些可能在他的公共言论或写作中，对他们的雇主或委托人，对他们的职业工作的赞助者，有某种潜在的利益冲突。

9.03 个人获得的避免

(a)在任何条件下，社会学家不能运用，甚至寻求从机密文本(如新写的手稿或一个建议性评论小组中得到的知识)获得机密，除非他们已授权你这样做或者这些信息已经可以作为公共知识。

(b)在任何条件下，社会学家不可以在没有得到雇主和委托人允许的情况下，从雇佣或委托交往关系的信息和材料中寻求获得。

9.04 工作场所中的决断

在他们的工作场所中，社会学家采取适宜的步骤去避免利益冲突的发生，详审可能的偏见取向和关系；在研究、教学和服务中，这样可能的偏见取向和关系包括，但并不限于：家庭条件、商业、亲密的个人关系，以及社会学家与之有强烈冲突和意见不一的那些人。

9.05 工作场所外的决断

在他们工作场所外的职业活动中，如果他们具有倾向性的隶属或关系，在任何条件下，社会学家应拒绝参与关于从个体和团体中分配和扣留报酬的商讨和决定。这些倾向性的隶属或关系如下：

(a)可以被认为从决定中受益的当前就业和被考虑在一个组织中就业。

(b)可能被认为从决定中受益的现任办公人员和董事会成员。

(c)可以被认为从决定中受益的当前就业和被考虑就业的同一个组织。

(d)个体可以从决定中受益的配偶关系、老乡关系或为人所知的其他关系。

(e)个体可以从决定中受益的商业和职业伙伴关系、研究合作者、雇员、上司和同学关系。

10. 公共交流

社会学家在进行关于职业服务、证明身份的文件、专长、工作产品和出版物的公共沟通中，应当坚持较高的职业标准，无论这些交流是来自他自己或别人。

10.01 公共交流

(a)社会学家要采取措施确保所有公共交流的精确性，这样的公共交流包括，但不限于：目录列表、私人简历、课程表、广告、印刷品的小册子、媒体的观点和

评论法律程序的申明、演讲和公共口语呈现或其他的出版物。

(b)社会学家不能因为他们的陈述、传达或建议,也不能因为他们的省略而做出与他们的研究、实践或其他工作活动或他们私人所联合形成的组织这样的活动有关的错误的、虚假的、欺骗的公共陈述。这些错误和欺骗性的言论涉及社会学家的下列活动:

- 训练、经验或能力
- 学术成就
- 证件
- 组织或机构友好关系
- 服务
- 酬金
- 出版物与研究成果。社会学家不应当发表与他们的职业服务中成功的基础、结果、程度相关的错误与欺骗性的言论。

(c)当社会学家根据公共演讲、证明材料、广播节目、录制的磁带、印刷文章、邮件材料或其他媒体而提供职业意见或评论时,他们应当采取预防措施确保:

- 这些言论基于合理的研究、文化和实践。
- 这些言论与伦理准则是一致的。

10.02　他人声明

(a)从事或雇佣他人发表关于推销他们的工作产品、职业服务的公共声明或从事其他活动的社会学家应当保留责任。

(b)社会学家应当做出努力防止那些不直接从事雇佣、监督的人(如雇主、出版者、发起人、组织成员、媒体成员)做出与他们的职业研究、教学和实践活动有关的欺骗性言论。

(c)在与新闻界、广播电台、电视台和其他交流媒体工作时或在媒体做广告时,社会学家要知道可能的利益冲突或冲突的出现(如他们不为媒体雇员提供补偿金),他们应当坚持高水平的职业诚信(如他们承认收到有偿广告)。

11. 机密性

社会学家有义务确保机密信息受保护,他们这样做是为了确保研究的完整和与研究参与者开诚布公地交流。为了保护在研究、教学、实践和服务中得到的信息,在收集机密信息时,社会学家应当考虑这些信息的长期使用,包括考察在公共档案中的位置和这些信息被其他研究者和参与者检验的情况。

11.01　保持机密

(a)社会学家采取预防措施,保护参与者、学生、雇员、委托人和其他人的保密权力。

(b)即使在没有法律保护和特权允许的情况下,社会学家仍然要对研究参与者、学生、委托人和其他人提供的机密信息保密,社会学家有义务保护机密信息,并且不能使通过机密手段得到的信息,以危害研究参与者、雇员、学生或其他人的安全或声誉的方式使用。

(c)在被理解为机密的情况下提供的信息直至提供信息者死亡也应被视为机密。

(d) 社会学家应当保持机密商议、活动、角色的完整性，这对包括职业委员会、评论专家组、建议团体(例如美国社会学协会职业伦理委员会)的活动，都是适用的。

(e) 在可能的程度上，在监督、建议文本中给出的学生记录、表现数据、私人信息，无论这些信息是口头的或文本的，社会学家都应保护这些信息的机密性。

(f) 保持机密性的义务延伸到能够获得信息的研究和训练团队成员与合作组织，确保获得机密信息受到限制，是研究者、管理者和督导人员的责任。

(g) 在使用由他或机构收集的个人隐私信息时，社会学家应当保护个人确认信息的机密性，当个体能合理要求信息不被用个人身份公开时，信息应当是私人的(如医疗和就业记录)。

11.02 保密限制

(a) 社会学家应当充分了解所有限制和改变机密的法律和规则。这些法律和规则赋予了社会学家确保完全机密的权利。相应地，在开始就应将所有确保机密的限制条件告知研究参与者、学生、雇员、委托人和其他人。这些限制与11.02(b)设立的伦理准则是一致的。

(b) 社会学家可能遇到不可预料的环境，他们意识到威胁研究参与者、学生、雇员、委托人和他人健康和生命的信息，在这种情况下，社会学家应当把机密确保的重要性与其他伦理准则、行为标准和现行法律平衡考虑。

(c) 在公共场所的观察，公开举行的各种活动，以及在没有法律和习俗约束隐私的情况下，机密是不必要的；同样，从公共记录中获得的信息也是不需要保密的。

11.03 讨论保密及其限制

(a) 当社会学家与人们建立了科学和职业的关系，他们讨论：

- 保密的相关限制。
- 通过他们的专业工作产生的信息带来的可预测效用。

(b) 除非是不可行的或反生产力的，有关保密的讨论在关系刚建立的时候就应举行，作为一种新环境这是正当的。

11.04 对信息的可能应用的期望

(a) 当研究需要在数据库和系统记录中留有个人身份信息时，社会学家在公开这些信息时应当删除这些个人身份信息。

(b) 如果涉及研究参与者、委托人和其他服务获得者的机密信息进入到数据库中或系统记录中，并且这些信息对人们是有用的，那么，在没有相关方面同意的情况下，社会学家应当剔除这些个人信息或者说采用其他技术手段不让个人信息显示。

(c) 如果删除这些个人信息不可行，社会学家应当在将这些数据传输给他人或评论他人收集的数据时，采取合理措施得到相关个体的同意。

11.05 机密信息的电子交易

当机密信息通过电脑网络传输和交流时，社会学家应当尤其谨慎。社会学家在用科技革新手段传输信息时，应当注意机密材料的保密问题，如通过公共电脑网络，可能

会使专业及科学的交流泄露给未被授权的他人。

11.06 来源的匿名性

(a)如果没有得到个人和法人代表的同意,社会学家不应当在他们的著作、演说和其他公共媒体中披露研究参与者、学生、团体或组织客户和服务获得者的个人确切信息和相关机密。

(b)当机密信息在科学或专业场所被应用时,社会学家应当隐藏研究参与者、学生个体和组织客户及服务获得者的身份。

11.07 减少对隐私的非法侵入

(a)为了减少对隐私的非法侵入,在进行公共交流时,社会学家在他们的书面或口头报告中,咨询和公共沟通中,只应包括与研究关系密切的信息。

(b)在讨论涉及研究参与者、学生、监督者、雇员及个体和团体委托人的机密信息时,社会学家应当仅仅为职业目的和仅仅与这些事情相关的人员讨论。

11.08 机密信息的保存

(a)社会学家应当采取合理措施确保记录、数据和信息在与伦理准则相符合的方式下保存,确认这些保存者受到法律或组织原则的约束。

(b)社会学家应当考虑在社会学家死亡、失去行动能力或者从岗位或实践中退出时,关于记录、数据和信息的机密性如何保持的问题。

(c)当社会学家将机密记录、数据和信息传输给他人或组织时,他们应当确保记录、数据和信息的获得者采取措施保护这些材料的机密性,至少应当像他们原来所承诺的那样。

12. 告知同意

告知同意是对人进行科学研究的最基本的伦理信条。如果没有他或者他的授权代表的同意,社会学家不能将人作客体进行研究,在这个准则中没有例外。社会学家应意识到不可预料的影响的可能性和由于研究者的研究专长和权威而给对象带来的些许压力,因而他们设计得到告知同意程序时应将这一点考虑进去。

12.01 告知同意的范围

(a)社会学家只有获得研究参与者或其法人代表的同意才能进行研究。

- 当要通过研究参与者的任何形式的交流、互动和干预中获得数据时。
- 当研究参与者的行为一般发生在个体期望没有他人观察和对其行为报告时。

(b)尽管同意具有极端重要性,在下列情况下,社会学家会选择放弃这个标准:

- 研究不会使研究参与者有任何危险。
- 如果要征求研究参与者的同意,研究就无法实施。社会学家应当意识到放弃同意原则,要征得制度性审核委员会(institutional review boards, IRB)的同意,如果没有这种委员会,就得征求其他擅长伦理研究的权威组织同意。在这种情况下,个人身份信息的机密性必须保持,除非不同于11.02(b)设立的准则。

(c)社会学家可在公共场所或应用现有的关于个体的公共信息(如,在公共场所的自然观察、公共记录分析,或文献研究)进行研究,这时不需获得同意。如

果,在这种情况下,社会学家对是否需要获得告知同意持有怀疑,在进行研究前,他们可以咨询制度性审核委员会,如没有这种委员会,应当咨询其他的具有研究伦理专家的权威组织。

(d)在对弱势人群(如,青年、新来移民、精神病患者)进行研究时,社会学家应当采取特殊措施确保这种研究的自愿性能被理解和认同,并且同意不是强迫的。在其他方面,社会学家应当坚持12.01所设立的原则。

(e)社会学家要熟悉并遵守政府和联邦关于获得同意进行研究的章程,以及制度性审核委员会的要求。

12.02 告知同意的过程

(a)如果需要告知同意,在进行研究前,社会学家应当就研究的实质和调查者的责任与研究参与者或者其法人代表达成一致。

(b)如果需要告知同意,社会学家应当对研究参与者或其法人代表使用能理解的语言,并尊敬他们。

(c)如果需要告知同意,社会学家应当让研究参与者或其法人代表有机会在研究的整个过程中就研究的任何问题提问。

(d)如果需要告知同意,社会学家应当告知研究参与者或其法人代表研究的实质是什么;他们应当保证参与者的参与及继续参与是自愿的;他们应当告知参与者影响他们自愿参与的因素(如他们参与的可能风险和利益);他们还应解释研究的其他方面,并对可能参与者的问题作出反应。而且,如果必要,社会学家还应当说明拒绝参与或在参与中退出不会受到处罚,并解释拒绝和退出的可预见的结果。社会学家还应当坦诚讨论机密性问题,以及按11.02(b)所设立的准则限定保密程度。

(e)如果需要告知同意,社会学家应当就口头同意做记录。他们应当意识到同意既可以是口头的,也可以是书面的。

(f)社会学家应当遵守对研究参与者所作出的所有承诺,并将之作为同意过程的一部分,除非出现不可预料的情况要求启用11.02(b)设立的准则。

12.03 学生和下属的告知同意

当在自己的机构或组织中用学生或下属做研究的时候,社会学家应当采取更加特殊的措施,保护可能的对象不会因拒绝和退出参与而对自身造成负面影响。

12.04 儿童的告知同意

(a)以儿童作为实验对象时,社会学家应在他们能提供这种同意的程度上,获得他们的同意。但,如果出现了12.01(b)的情况时,不需要同意。

(b)以儿童作为实验对象时,社会学家要获得其父母或法定监护人的同意。在下列情况下,可以不需要其父母或法定监护人的同意:①研究几乎不会对参与者造成任何危险;②如果需要同意,研究无法进行;③为了保护这个儿童(比如弃童和受虐儿童),其父母或法定监护人的同意并不是一种必然条件。

(c)社会学家应当意识到,放弃儿童及其父母或法定监护人的同意需要制度性审核委员会的同意,如没有这种委员会,应当咨询其他的具有研究伦理专家的

权威组织。在这种情况下，个人身份信息的机密性必须保持，除非遇到11.02(b)准则涉及的相关情况。

12.05 研究中欺骗的使用

(a)社会学家不能使用欺骗技术，除非：

- 欺骗的运用不会伤害研究参与者；欺骗的价值被可能的科学的、教育的、应用的价值所证明；如果不用欺骗的方法，就没有可行的被选择程序。
- 已经获得制度性审核委员会的同意，如果没有这种组织，就要获得其他的具有研究伦理专家的权威组织的同意。

(b)社会学家不能在影响参与者参与意愿的方面欺骗参与者，比如可能的生理风险、不舒适感、不良情绪体验等。

(c)如果欺骗是研究设计和实施的不可分割的一个特点，社会学家应当纠正研究参与者后来关于研究结论的错误概念。

(d)在极个别的情况下，社会学家需要隐瞒他们的研究者身份，比如，如果被知道，研究就无法进行。在这样的情况下，如果研究涉及研究参与者很小的风险，并且制度性审核委员会同意用这种方式进行实验(如没有这种委员会，应当咨询其他的具有研究伦理专家的权威组织)，社会学家可以隐瞒其研究者身份。在这种情况下，个人身份信息的机密性必须保持，除非遇到11.02(b)准则涉及的相关情况。

12.06 记录技术的运用

社会学家应在获得研究参与者、学生、雇员、委托人和其他人的同意后，才能录音、摄像和进行其他形式的记录，除非这些活动是在公共场所的自然观察，并且保证这些记录能在不伤害个人身份的前提下被运用。

13. 研究的计划、执行、传播

社会学家有义务促进研究的完整性，并确保在研究的计划、执行、传播中遵守科学的伦理信条。这样做是为了积累知识，减少结果误导的可能性，并保护研究参与者的权利。

13.01 计划与执行

(a)在研究的计划与执行阶段，社会学家要减少结果误导的可能性。

(b)社会学家要采取措施对研究参与者和其他受研究影响的人进行保护。

(c)在他们的研究当中，社会学家不应当使研究活动或自身行为以对研究参与者和他人的健康和生命有害的方式进行。

(d)在研究的计划和执行中，社会学家应当向专家咨询被调查的或可能受到影响的特殊人群的有关问题。

(e)在计划和执行研究中，社会学家要按照伦理准则设立的原则考虑其可接受性。如果最好的道德实践是不明晰的，社会学家应当向制度性审核委员会咨询，如果没有这种组织，应当向其他的具有研究伦理专家的权威组织咨询。

(f)在由他们监管或控制的研究中，社会学家应当对他们自己或别人实施的道德行为研究负责。

13.02　不可预期的研究机遇

如果社会学家在教学、实践、服务或进行非职业化的行动中，决定他们要去从事以前未曾预料的研究，他们应知道他们自己的意图，并采取一些措施确保他们所从事的研究能与道德准则协调，特别是那些与机密及告知同意相关的准则。在这种情况下，社会学家应当获得制度性审核委员会的同意，如果没有这种组织，就要获得其他的具有研究伦理专家的权威组织的同意。

13.03　为研究参与者提供诱因

社会学家不能通过提供过分的、不恰当的经济的或另外的一些诱因去获得研究参与者的参与，特别是强迫参与者参与。但是社会学家可以向研究参与者提供一部分动机，如可得到的和适合的资源作为诱因。

13.04　研究的报道

(a)社会学家可以传播他们的研究发现，除非碰到未曾预料的情况(如，研究者的身体健康状况)，或者与雇主、签约人、代理委托人签定的协议禁止这种传播。

(b)社会学家不能捏造或在他们的出版刊物中虚构结论。

(c)在陈述他们的著作时，社会学家应当完全报告他们的研究成果，不能删减有关数据，不管他们支持或者是反对所期望的结果，他们都应如实报道。

(d)社会学家应特别严谨地陈述研究发现的有关条件和限制，以及他们研究的说明，社会学家也应该提出一些与他们著作阐述有关的潜在推测、理论、方法和研究设计。

(e)要与自己在整个研究发现中采用的方法和分析所秉承的精神相一致，一旦自己的研究发现被广泛公开传播，社会学家应允许一些有责任感的研究人员对他们的研究进行公开的评价和确证，研究人员要用合适的安全措施来保护研究参与者的隐私。

(f)如果社会学家在他们的数据中发现严重错误，他们就应采取合理措施，如用更正、回收、出版勘误表等形式纠正错误。

(g)社会学家应当用书面材料报告经济支持的来源以及与发起人的任何特殊关系。在特殊情况下，如果他们已对发起人的性质和利益进行了充足的描述，社会学家可以隐去特殊发起人的名字。

(h)在为他人提供教学、出版、实践和服务的工作时，社会学家应当尤其注意采用正确的信息和标准报告他人的学术成就。

13.05　数据共享

(a)作为常规研究，社会学家应共享相关的数据和文献。社会学家在完成其研究项目和他的主要出版物后，应当公布数据，除非与雇主、合同委托人先前签署了这方面的协议，或者是因无法共享数据或为了保护数据的机密性和研究参与者的匿名性(如人种学的原初记录和详细信息)。

(b)无论数据共享是否可行，社会学家应期望将数据共享作为研究计划的一个整体。

(c)社会学家应当在与研究参与者的利益一致和保护他们提供的信息的机密性

的前提下共享数据。无论法律要求与否,他们都应保持数据的机密性;在数据共享前,要去掉个人身份信息;如果必要,要采取其他避免措施。

(d)在研究结果出版和传播后,社会学家可以在一段合理时期内保存有用的数据和与研究相关的文献,而不必提交给公共档案馆。

(e)如果必要,社会学家可以要求索取数据进行尝试分析的人支付相关费用。

(f)社会学家从其他人那里得到数据进行尝试分析,应当意识到起始研究者的贡献。

14. 剽窃行为

(a)在出版、教学、实践中,社会学家应当确认、指出他们所引用的数据和原材料的作者,而无论它们是发表的,没有发表的,还是电子文档的。

(b)在出版、教学、实践中,社会学家引用他人著作的知识,或作为参考书,即使不是逐字引用和成段引用,他们也不应将他人著作视为已有,而无论被引用的是发表的,没有发表的,还是电子文档的。

15. 原著名誉

(a)社会学家仅仅为他们自己所贡献的部分承担责任和享受名誉,包括原著名誉。

(b)社会学家保证基本原著名誉和出版名誉是基于相关个体的科学工作者和职业贡献,而无论他们的地位如何。在声明和决定作者顺序时,社会学家应当仔细考虑在研究和写作过程中的主要参与者的贡献。

(c)如果出版物主要来自学生的论文,在多作者的情况下,学生通常作为第一作者。

16. 出版过程

如果社会学家是作者或编辑,他们在出版和审核过程中应当坚持高伦理标准。

16.01 投稿

(a)在多作者情况下,社会学家应当在投稿前与其他作者商谈,并达成互相可接受的一份协议。

(b)在投稿时,社会学家应当保证出版物是第一次投稿,除非有明确的政策允许多次投稿。社会学家不应将原稿交给第二家出版机构,除非得到第一家出版社的正式决定或原稿被退稿。社会学家在将原稿交给杂志社、出版社和编辑部审核时,在他们作出正式的接受决定前,可以退稿。

(c)社会学家可以将书的原稿交给多家出版商。然而,一旦社会学家签定了合同,他们就不能再退稿,除非有合理的原因这样做。

16.02 数据的重复出版

如果社会学家出版的数据或结论在先前已经出版过,他们在出版时应当给予恰当的声明。

16.03 编辑的责任

(a)如果作为杂志和丛书的编辑,社会学家应公正运用标准,在操作时不应有个人和意识形态上的偏袒和恶意。作为编辑,社会学家应当能认识到可能的利

益冲突。

(b)如果作为杂志和丛书的编辑,社会学家应当与保持机密的实践一致,确保审核过程的机密性。

(c)如果作为杂志和丛书的编辑,社会学家有义务出版编辑部接受的所有原稿,除非在接受后发现了重大错误和严重的伦理越轨问题(如抄袭和科学误导)。

(d)如果作为杂志和丛书的编辑,社会学家应当确保审核成员的匿名性,除非他们得到了审核人员的同意透露身份。编辑应当保证其职员遵守这个实践。

(e)如果作为杂志和丛书的编辑,社会学家要保证作者的匿名性,除非原稿已经被出版商接受或者杂志社的业务是其他的行为。

(f)如果作为杂志和丛书的编辑,社会学家应当为所有原稿提供及时的审核并对审核情况做及时反应。

17. 审核者的责任

(a)在审核为出版提交的材料、保证书和其他评价材料时,社会学家应当顾及过程的机密性和提交材料人的所有者权力。

(b)社会学家应当提示利益冲突并拒绝涉及利益冲突的他人著作审核的要求。

(c)当他们相信审核过程会有偏见或对整个过程有疑问的时候,社会学家可以拒绝对他们著作进行审核的要求。

(d)如果被要求审核他们已经审核过的原稿、书本和计划,社会学家应当让要求者知道这些已经审核过(如编辑、项目官员),除非他们明确要求进行重新审核。

18. 教育、教学和训练

作为教师、监督员和训练专员,社会学家应当坚持较高伦理准则,以确保社会学家教育的质量和教师—学生关系的完整性。

18.01 教育计划的管理

(a)负责教育和训练计划的社会学家,应当保证所有教育计划都是正当设计的,能提供适当的经验和达到计划所要达到的目标。

(b)负责教育和训练计划的社会学家,应当保证对计划内容、训练目标、训练对象,以及必须符合满意完成计划的要求有一个完整描述。

(c)负责教育和训练计划的社会学家,应采取措施保证助手和临时教员有坚实的知识和方便学生学习所必需的教育技能。

(d)负责教育和训练计划的社会学家,有义务保证授予毕业生的道德成为他们职业准备的一个部分。

18.02 教学和训练

(a)社会学家应当有意识地履行他们的教育责任。他们应当有适当的技能知识,并接受了适当的训练。

(b)社会学家在课程开始就应提供充足的信息,尤其是关于研究对象所涉及的范围、评价的基础和课程经验的本质。

(c)社会学家应不受制于经济和其他诱因,对教材、课程内容、课程要求和以教育

标准为基础的排名作出决定。

(d)社会学家应当对他们的助教和其他教育培训者进行培训和监督,并采取措施保证这些人能负责的、胜任的和合乎道德的履行其责任。

(e)社会学家不能因为自己对同事怀有敌意或智力不如同事,而阻止学生和被监督者与这些同事交往,社会学家也不能干预学生和被监督者的学习、学术成长和职业发展。

19. 契约性和咨询服务

(a)社会学家只有了解他们计划使用的实物、方法和技术或打算与合适专家合作时,他们才能着手承诺、合同和咨询。

(b)在着手承诺、合同和咨询时,社会学家只能基于适当的信息和技术形成他们的职业工作的结论。

(c)在承诺、合同和咨询时,已经得到了项目的经济支持,社会学家应当努力按日程完成预期工作。

(d)在着手承诺、合同和咨询时,社会学家应当用文件精确证明和适当保留他们的职业和科学工作。

(e)如果已经建立了研究、咨询和其他服务的安排,社会学家应当在开始就在可能的程度上澄清与个体、组织和组织客户关系的本质。相应地,这种澄清包括履行服务的本质,所提供的服务的可能运用,为学术和出版目的社会学家将来运用这些著作的可能性,交付这些服务的时间表,以及赔偿和广告安排。

20. 坚持伦理准则

根据伦理准则,社会学家有义务面对、谈论和尽力解决伦理问题。

20.01 熟悉伦理准则

社会学家有责任去了解伦理准则、其他应用的伦理规范及其在自己工作中的应用。没有意识到或误解一个伦理准则本身不是对不道德行为的掩盖。

20.02 正视伦理问题

(a)当社会学家不确定一个特定的环境或行动过程是否违背伦理准则时,他们应当向其他了解伦理问题的社会学家咨询,或者向美国社会学协会的职业伦理委员会 ,或者向制度性审核委员会这样的组织实体咨询。

(b)当社会学家采取行动或面对在伦理准则中阐明的伦理标准与法律要求的冲突时,他们应当将之告知这些准则委员会,通过同事、职业组织或向美国社会学协会的职业伦理委员会咨询,采取负责任的方式解决这些冲突。

20.03 在伦理纠纷中公平对待各方

(a)社会学家不应因某人进行伦理投诉而歧视他。

(b)社会学家也不应因某人是伦理投诉的对象而歧视他。这并不妨碍基于伦理投诉的解决结果而采取行动。

20.04 报告他人的伦理越轨行为

当社会学家有充分的理由认为另一个社会学家存在伦理越轨行为时,且非正式解决方案是合适和可能的,他们应当通过引起个体注意的方式尝试解决这个问题;假定

这种行为不违背任何保密权力，根据这种理念他们应寻求是否和如何继续进行研究。这种行为包括上文提及的美国社会学协会的职业伦理委员会的例子。

20.05　与伦理协会的合作

社会学家在伦理调查、伦理推进和因美国社会学会的要求而进行广泛合作。这样做，他们可以努力解决任何保密问题。而不合作将被认为是伦理越轨。

20.06　不正确的投诉

社会学家不宣扬或鼓励那些倾向于打击被认为是伦理越轨者的利益，而不是出于保护程序的公正和公众利益的伦理投诉案。

注：这个伦理准则在1997年春季经过了美国社会学协会职业伦理委员会的通过。使用获得了授权。

3.8.2　制度性审核过程

像常规行为一样，从联邦机构得到资金的学术或研究机构（大部分是领导机构或几乎所有大学和学院）都是用人作对象从事实验研究。这些审核由制度性委员会进行，其成员由教员、职员、学生代表和管理者联合组成。通常地，无论谁从事研究（研究生、本科生，或者职员），都需要呈交获得同意的申请。

例如，在明尼苏达大学，Mankato 人类对象保护制度性审核委员会是一个建立在大学中的委员会，它由教员、管理者、健康职业工作者和社区成员组成。制度性审核委员会保护人类研究对象和参与者的权力和福利。制度性审核委员会的管理者是负责提供制度性审核委员会所需要的资源和支持的人。在明尼苏达大学，研究生学习和研究的主任是制度性审核委员会的管理者。

下面是制度性审核委员会在堪萨斯大学颁布的一个文本的副本，在那里，制度性审核委员会被称为人类实验咨询委员会（Advisory committee on Human Experimentation，ACHE 文件要经过制度性审核委员会的同意才能再印），这些文件为那些没有很好组织的审核实践提供了一个很好的模型。

堪萨斯大学制度性审核委员会文件

向堪萨斯大学人类实验咨询委员会呈交议案的说明书
（*ACHE*）

Lawrence ，KS 66045-7563

谁必须请求 ACHE 的同意以及为什么要征求其同意

1974/1983（PL93—348）国际研究法案设立的指导原则是，为了使组织能从联邦资源中获得行为和生物医学研究的同意，IRB 必须成立，并审核涉及人类客体的研究。IRB 的义务是保护作为研究对象参与这些机构的研究的人的权利。堪萨斯大学的

IRB 是 ACHE,大学必须获得关于大学、研究者和 IRB 关于人类客体研究责任多重项目担保的同意,我们的多重项目担保可以在 http://www. research. ukans. edu 上找到。ACHE 评论涉及以人作研究对象的研究的建议,如果研究:

- 是大学自己发起的(如与大学联合的)。
- 作为研究责任的一部分是由大学雇员或办事机构执行或领导的。
- 使用大学教员或财产,并由大学雇员或办事机构执行或领导的。
- 涉及应用组织非公共资源去联系和确认参与者或可能的参与者。

以人作为研究对象的调查者,并且这些对象符合上述任一条件,他们就得准备呈交给 ACHE 获得其同意的计划。ACHE 对计划的评论遵循联邦规则章程(Title 45, Part 46),其中设立对人类客体保护的最低标准。

什么时候上交计划

申请 ACHE 同意的计划可以在学术年度的任何时候呈交,ACHE 应当对计划进行不间断审核,通常在第一周就要进行审核。在每个月的第 15 天收到的研究计划会在下月的会议上讨论。

例如,如需要全体审核委员参加的计划在 10 月 15 日呈交的,委员会决定将会在 11 月作出。而在 10 月 16 日提交的计划直到 12 月委员会才能确保作出决定。那样的话,委员会审核计划工作所需的时日至少是 3 个周或 6 个周。如果 15 号恰好在周末或假期,计划接收最后期限是下个工作日的下午 5 点。

ACHE 提倡获得委员会同意的时间应当计算在研究的日程表里,尤其是当委员会行动(如关于项目计划的讨论和投票)不等于是委员会的同意和保证。而且,同样的往返所需日期不能在夏季保证,因为一些委员会成员正在履行 9 个月的大学合同,或者他们在参加暑期会议,那样就不能参加审核。

提交什么

采用 ACHE 的申请表格提交计划。样表可以从 ACHE 办公室或你们系得到,他们也可发邮件给你。

最新申请版本在首页的左上角打印有“1/98”,提交给 ACHE 的申请表包括两部分。

第一部分:ACHE 申请表格包括四页(首页、审核页、描述页、摘要页)。手写文本后应当有申请的副本。IBM 或 Macintosh 申请版本的申请字处理器的安排在 ACHE 办公室仍然可得到。给 ACHE 办公室寄张磁盘,我们将形式要求拷在磁盘后归还给你。ACHE 可以通过邮件对你的要求发出说明书和申请表格。

第二部分:申请书的附录也应当一起提交。附录包含申请书的基本信息,通常包括(但不限于),执行研究的书面同意的副本,其他附录材料的副本(调查、问卷、评价材料),这些在研究过程中将会用到的其他材料的副本。然而,PI 也可包括那些他们认为在委员会评价过程中有用的材料。

ACHE 计划中这两部分的任一部分在接下来的过程都会讨论。

第一部分:ACHE 申请书

书后附有 ACHE 申请书。进行新的提交时请用以下形式:

(第 1 页)首页包括 ACHE 申请书的常规形式。首页右上角的 ACHE 申请书号的地方应该是空白的。

1. 调查者的名字。所有的调查者应当列出。多于一个调查者的研究项目,应当署首席调查者(principal investigator, PI)的联系方式,PI 可以是堪萨斯大学的职员、学生或教员。如果 PI 既不是教员也不是职员,堪萨斯大学的职员督导或发起人应当在表格上签名(参照第 5 项)。因此,学生需要有职员督导或发起人。职员督导或发起人在项目的适当执行中对人类客体保护最终负责。

指导:所有的研究者,包括项目职员督导,在项目获得 ACHE 的同意前,必须就人类客体研究完成任务执行研究的指导。你可以在 www. research. ukans. edu/fufor 上获得。

2. 联合部门。如果申请,请指明 PI 属于学校哪个部门。

3. 大学或家庭邮政地址。这是关于计划的通信要发送的地址。如果可能,可以给出大学地址,以方便 ACHE 可以用大学邮件发送信件。如果你有 E-mail 地址,你应当提供,计划审核的日期用 E-mail 比平信更快。

4. 电话号码。请提供 PI 能收到的大学电话号码和家庭电话号码。这可以使协调者能就申请书的技术问题与 PI 保持联系。

5. 负责项目的职业成员的名字。提交给 ACHE 的研究项目必须有发起人。如果 PI 不是堪萨斯大学的职员,请指明谁是研究发起人。

6. 调查的类型和活动的性质。如果申请 ACHE 同意的 PI 是堪萨斯大学的职员,这个项目的上层应该填出 PI 的地位,研究是内部还是外部资金。如果堪萨斯大学的资金已经分配给了这个项目,请标明数目。如果 PI 是学生,在项目的底端标明。

7. 调查的题目。给出项目的简洁主题。简洁的题目方便通信和做记录。如果这是一个资金项目或一个调查题目,请给项目一个简短标题。简短标题能促进交流和档案保存。如果这是一个基金项目或者是一个正在申请基金的项目,且和 ACHE 项目标题不一样,也请给出标题。

8. 如果申请个人不是堪萨斯大学的教员、职员,或者学生,必须检查其材料,并且指出其他可能参与这个课题研究的非堪萨斯州大学人员的姓名。

9. 各种证书。PI 和任何参与研究的人员都应该在首页底部签名,从而确认大家都熟悉 ACHE 的有关政策和尊重被试这种伦理行为的核心理念。这是非常重要的一步。如果没有项目主持者和所有课题参与者的签名,ACHE 将不会接受申请。请注意,通过 E-mail 或者硬拷贝的方式提供申请时,一定要能证明你已经阅读、弄懂,并将遵从堪萨斯大学的政策和程序。这些政策和程序是关于保护研究中被试的标准,以及申请人赞成这些标准并坚守 ACHE 的政策和程序要求。

(第 2 页)请在该页顶部填写 PI 的名字和正确的标题。下面标有“ACHE#.”的地

方是给官方用的,要保证空白。

10. 清单。清单允许 ACHE 迅速地筛选那些可能加速或免除条件的申请。请仔细阅读条款 10a 到 10m,并回答“是”或者“不是”。如果关于合作机制的条款 10c 标注为“是”,那么就要提供“其他支持材料”。如果条款 10k 标注为“是”,那么研究工具的副本(或者对工具进行详细的描述)必须以附件的形式附上。这些物件必须保存在与此项目有关的 ACHE 档案中。

声像档案的运用

如果项目中要运用声音和(或)图像档案,你的摘要和内容应该解释谁将有可能接触这些磁带,你将采取什么安全措施来保护被试资料的隐私,以及项目完成后如何处置这些磁带(例如消磁、保留下来以用作将来研究等措施)。

支付被试的报酬

如果研究中包括给被试支付报酬(也就是说,如果条款 10b 标注为“是”),请注明这种支付经过了仔细考虑,是合理的。这种报酬只能是打算补偿作为研究参与者的被试在交通、时间上等的付出,不能用可能潜在强制的付款计划。

11. 大约有多少被试应该包括在研究中。请简要说明你计划从多少被试处收集数据。

(第 3 页)有关课题目标、选中的被试和挑选程序的信息,将有助于委员会洞悉研究带来的潜在利益,被提议的人口是否特别易受攻击或处在危险之中,以及选择被试的过程是否公平,是否充分考虑到机密和隐私。

12. 课题目标。请在申请中对课题目标进行简要描述,同时不要用行话。请只限于在提供的空间内填写。

13. 描述选中的被试。请简要说明对项目涉及被试选取或排除的任何特殊标准。例如,被试被包括在项目中只是因为他们来自一个特殊的年龄群体、种族群体,或者是性别,请在此做简要说明。另外,在研究中,如果包括那些具有某些医学属性(例如 Alzheimer 病,心脏病等),或者身体属性(例如跑马拉松的人、骑自行车的人和举重者)的被试,也请在此做简要说明。

选择被试的注意事项

14. 要说明如何选择被试。请指出你将如何与被试接触并招募他们参加研究。譬如,如果被试是通过一个学校内的被试库,或者通过心理学 104 集中放映室,或者通过一个合作机构随意招募(用邮件或电话),都请在此详细说明。如果想要从一个组织(例如诊所、医院)的病人中招募被试,调查员应该确认被试的利益,而不是在没有征得病人的同意就从这些组织中获得病人清单。调查员可以请求组织分发同意书或者介绍信,以及让相关的被试直接与调查员联系或者通过组织联系。这种方法用来保护被试的隐私和预期。

(第 4 页)这是申请中重要的一部分。虽然第 2 ~ 4 页的提议在月会之前提前分发给委员会的每个成员,但只有 ACHE 的三个成员全部阅读和考虑这些提议。因此,摘要将被委员会所有成员看到,那么它应该完备而且对提出的项目提供准确的陈述。

委员会评价提议过程中遇到的许多问题在于对提议内容理解困难。通常委员会要求在对任何提议作出最终决定之前要求保持头脑清醒。通过清晰的摘要和常用语言,许多问题就能得以避免。

15. 程序摘要。摘要应该简洁地对整个项目作出总结。要描述项目涉及的所有程序。ACHE 的委员会由来自种种学术或非学术领域的专家组成。因此,根据这样的对象能够理解的方式来描述程序是非常重要的。只在提供的空间内来填写程序,同时不用行话、不缩写或者不用技术术语。如果你必须用技术术语,请对此进行详细说明或者作出解释,以利于那些没有此领域专业背景的人能够看懂。

第二部分:附　录

附加物应该以附录的形式附在申请书后。这些材料通常包括下列部分。

研究工具的复制品

通常来说,如果申请书第 2 页的条款 10k 被查明为“是”,那么在提议的研究中将要运用的工具(估价法则、测量、问卷、调查、协议等等)的复制品要放在附录中。如果工具或协议不能作为附件,那么就应该对工具进行足够详细的描述。

告知同意表格

如果同意使用了告知同意的内容,附录中应该有告知同意表格的复制品。根据联邦规则章程(45 CFR 46),同意表格必须包括以下条款才算合适和可以使用。

1. 陈述研究目的和简要描述即将进行的研究步骤。

2. 在相当程度上能预见的关于被试的任何不适或危险(心理上的、社会上的,或身体的)要进行描述。如果项目中不只包括最小程度的危险,那么就应发表关于堪萨斯州民事侵权主张行为的声明。最小程度的危险是指研究中预料的危险或不适的可能性,与程度不会高于这些在日常生活中遇到的或者是在常规的身体或心理检查或测试过程中所表现的。堪萨斯州民事侵权主张行为的主要内容如下:

> 在受到伤害的情况下,如果能够证明这种伤害是政府雇员在其工作范围内由于疏忽、错误行为,或者是失职而引起,堪萨斯民事侵权方案将主张补偿。

3. 陈述从研究中能够相当程度地期望给被试或其他人支付的利益。

4. 揭发二者择一的程序将对被试有利(这种情况通常只适用于实验医学治疗的研究)。

5. 有参与自愿的表达。这应包括随时可以终止参与,拒绝参与,或者退出参与的保证,以及项目保证被试其他方面的权利没有处罚或者利益的损失。

6. 陈述可以确认的被试记录的机密性将被保护的程度。

7. 及时指出研究中作为参与者所要承担的义务。

8. 满足有关项目的任何要求。在数据采集阶段完成后,万一有问题出现,及时通知相关人员。这经常是通过提供 PI 名单、部门联系,以及同意表格书上的电话号码。

9. 签名

- 如果合适(例如,被试在法定年龄之内,有能力弄懂并提供告知同意书),则要求被试的签名。如果不满法定年龄(18岁以下)要求,个人可能会无意中被招募来参与项目的研究(例如,当被试是从大学被试库获得的时候),在其签名处下方应该注明,“以我的签名,证明我至少有18岁”。
- 如果被试的签名不合适,则要求其父母或监护人的签名(必须在法定年龄之内)以及被试的名字。

10. 如果在同意表格书上没有出现给被试的说明,那么就应该指导被试如何理解接下来的任何说明。

11. 主张在签名下方注明,“通过签名我承认我已经获得了同意表格书的复制件并保存”。

12. 在同意表格书或者介绍信上添加该系管理者的名字、部门和部门电话。

请注意,如果申请表格书超过1页,那么每页的页码应该能充分表明整个表格的长度(例如,“第1页,共3页”,“第2页,共3页”,等等)。已经收到的同意表格书的复制件规定了说明性的目的。

如果计划使用口头通知的同意,PI应该考虑方向和规章证明口头同意在联合核心规则(45 CFR 46.117.b.2)内。它特别包括服从文本或者履行同意手续的手迹,证人到场,以及此研究为被试提供的概括描述。

同意程序

在研究中对儿童或者只能采取通知同意的其他参与者,则采用其他的折衷方案,通常适宜的方法是在数据采集阶段获得他们某种形式上的同意或者“赞成”。比如,尽管儿童或者残疾人不能提供参与研究的告知同意,研究者仍然应该用被试能够懂得的语言解释整个程序,并获得他们口头上的“同意参加”。如果同意程序被启用,这个程序的“原本”的原型应该包含在申请的附录中。

调查和问卷

邮寄调查或问卷本质上要求完全匿名,通常弃用签署告知同意函的方式。但是,在这种情况下,在通知同意的地方应该随调查附上“介绍信”。介绍信应该包括告知同意表格的危急方面(比如,利益风险状态、自愿参与的保证和回答的机密,等等)。同时还应包括问卷回收后回答者关于他(她)同意参与研究的陈述。介绍信的复制品应列入附录中。

如果调查者希望利用互联网或者电子邮件来进行调查,必须进行一些特别的防范。因为当回答者通过E-mail返回调查结果时,他们的电子邮件地址也随之而来。PI应该制订方案来消除这些信息,从而获得回答者的机密和匿名。同时,无论有意或意外,不仅是有意的接收者,其他人也有可能看到被试的答案,因此调查者应该告知被试,尽管作出了保护他们隐私的各种努力,但参与者的回答的安全性与机密性难以保障。

其他支持物件

一些委员会的关注点通常表现在被支持信的内容中。这些支持信来自研究中或与研究合作的机构(学校、诊所、关心健康机构、法律强制执行部门,等等)负责的个人。这些信可能提供迹象,表明他们在研究行为中合作的意愿,或者他们同意这样的机构中项目职员有权调查被试人群。这些信还暗示,在那种情况下,这样的个人已经回顾或者(如果已被运用)已经获得研究协议的安全承诺。

提交物品

调查者必须提交的东西有:已经完成的整个 ACHE 申请表的复制品一份(1~4页),同意表格书和所有附件材料。

提交地点

如果通过美国邮政提交建议,寄往:David Hann 先生,ACHE 协调者,233 Youngberg Hall, University of Kansas, Lawrence, KS 66045-7563。E-mail 申请书可以提交。学生研究者应该将他们的申请用 E-mail 发至他们所在系的领导,由他们审阅后再寄至:dhann@ukans.edu。请注意,那些不能用 E-mail 寄送的支持物件,必须亲手递送或者用普通邮件寄送。

校园邮件可以填写以下地址寄送:David Hann,233 Youngberg Hall。建议可以直接送至在 233 Youngberg Hall 办公的 Hann 先生。他的电话号码是(785)864-7429。不要把 ACHE 要求的提交物送给委员会的主席,这将导致延误建议审阅的进程。

评价 ACHE 建议的基点

保护被试的综合论点

在委员会被赋予了基本职责的基础上,ACHE 关注的建议存在一些特殊的方面。在法律上,ACHE 只能正式批准那些能够满足必备条件的计划。预先评价这些必备条件并在建议中对它们充分关注的 PIs,在 ACHE 审议过程中会明显遇到少得多的问题。审议的必备条件(与联合核心规则的解释类似)包括以下要素:

被试的危险降到最低限度。这是 ACHE 在浏览申请书时首要的,也是最重要的关注点。如果参与该项目,将招致什么样的潜在危险、压力,或者不适(如果有)? PI 在设计或者实施研究过程中,采取措施来减少这些危险或不适发生的可能性了吗?

被试的危险与期望利益之间相关联。如果有,能确保被试从此研究中获得的利益超过可能带来的危险吗?

告知同意书将能从每一位被试(或者是被试的法定代表)和文件中得以体现。被试完全被告知参与研究的危险和利益了吗? 被试被告知他们在参与过程中的基本权利(例如,没有伤害地退出)了吗? 告知同意的所有合适内容都包括了吗? 如果没有,这些省略证明是正当的吗? 如果被试自己不能提供告知同意,那么被试的同意是从与他们相关的人那里获得的吗? 能提供这个过程吗? PI 提供的关于告知同意方面的文件充分吗? 请指出 ACHE 在同意项目时所遇到困难的最常见形式,这些困难包括告知同意表格上的技术问题(比如,同意表格要求所列出的必要方面并没有包括进去)。

在管理数据上,有着充分的防备措施以此保障被试的安全。对研究予以监控从而

让PI关注先前无法预料的危险吗？这部分也包括被试个体在医学的或心理的危险被确认时的数据管理。举例来说，在心理测试实验过程中，一位PI可能发现被试心杂音的症状；在对压力的心理测试过程中，一位PI可能发现被试处在某种自杀倾向的危险中。如果出现类似情况，对研究过程中这种危险行为，有什么防备措施来保护被确认的被试？

充分的防备措施来保护被试的隐私并保守机密。关于被试反应的机密和匿名性保护达到了什么程度？被试的姓名将与他们的数据联系在一起吗？谁有权通过分析材料（比如，数据单、录音或者录像带）来确认被试？反应记录将被保存在安全的地方吗？在研究完成后，将有什么计划来处理那些可以通过分析而确认被试的材料？有适当附加的防护措施来特别保护那些有弱点的被试吗？

包括欺骗主题实验的ACHE政策

ACHE认识到为了保证调查的科学完整性，在研究设计中偶尔采用欺骗是必要的。然而，因为研究参与者关于研究的真实目的和程序的了解是被故意误导的，ACHE认为这种研究没有符合联邦规则章程（45CFR. 116. a. 1）所阐明的关于告知同意的基本要求。法律的这部分描述了告知同意的基本要素，而且声明在寻求告知同意时，下面的信息应当让每个对象了解：

……关于研究目的的解释……所实施的程序的一个大概描述……

请注明联邦法律不必然限制“告知同意”的概念和对象在研究开始时签订的同意的内容。

然而，联邦规则章程确实提供了告知同意可以改变和放弃的例子。在这种联邦法律下（45CFR. 46. 116. d. 1-4），这种情况只有在下列条件满足时才可以发生：

- 研究对被试不会有任何小的风险。
- 放弃和改变（同意）不会影响对象的权力和福利。
- 如果不放弃和改变（同意），实验就无法进行。
- 无论在什么适当的时候，对象在参与后可以了解相关的信息。

因此，向ACHE提交计划的调查者，在数据收集过程中误导参与者对研究目的和程序的了解，必须解释他们的过程是如何符合这些条件的。这种要求应当可以在许多方面得到满足。然而为了能陈述这些问题和方便这些建议的审核，ACHE推荐涉及欺骗的研究建议应当包括：

1）欺骗的正当化。正当化应当符合上述的第3项条件。调查者应当提供特殊和使人相信的理由，解释为什么完全的告知同意对本研究不合适，和完全的告知同意在何种程度上有碍于研究的完整性。

2）没有或极小风险的明确声明。这条声明符合上页列出的第一和第二条。调查者应当对不会对研究参与者有任何小风险作出明确的保证声明。联邦法律（45CFR 46. 102. g）对“极小风险”定义如下：

"极小风险"意味在计划的研究中所预期的伤害的风险不大,从可能性和强度的角度看,要比日常生活中或一般的生理或心理测试中发生风险的可能性要小。

3)盘问详情的描述。盘问详情的程序应当符合上列第4款条件。实验对象应当被告知存在欺骗,也应当被告知研究的真正目的何在,以及欺骗在保证研究的完整性中的作用。他们还应当被告知他们有现在就退出研究的权力。这可以通过一系列的环节来完成,包括盘问详情影响的简单说明,以及让参与者在研究结束时签署一份再次同意的表格。

联邦规范准则的文本中注明,在盘问详情被裁定为"不合适的"的时候,情况就有可能发生。ACHE允许这种可能性,但是PIs指出,在研究中需要被强烈证明为正当的盘问详情的要求可以省略。

尽管涉及欺骗的建议必须符合告知同意变化的条件,作出标记的告知同意仍然需要符合45CFR 116条款要求。A.2.8所描述的要求,它指明告知同意形式不应包括误导行为,也不应当是欺骗行为的一部分,委员会命令协调者不要呈送不完整的申请材料给他们。因此,呈交包含欺骗的研究申请,而又没有充分盘问详情的程序的申请者,要求在将申请材料呈交给委员会审核前,呈交给他们。

审核程序

每年度ACHE收到超过500份的研究建议。大约40%~50%的项目在每学期初收到,也就是在每学期的头两个月,ACHE要处理超过100个申请,余下的申请将在本年度的其他时间处理。在ACHE收到申请后,他们就开始进行审核处理程序,这个程序在由大学与美国健康与人类服务部所颁布的保险声明中得到了详尽的阐述。

免除责任与可以迅速完成的项目的筛选

ACHE协调者或者ACHE主席首先筛选项目以决定他们是否符合可免除IRB项目审核项目的描述,或者它们应该进行一次迅速审核。

免除审核。如果项目如联邦规定所规定的,符合下列一项或几项要求,可免除全体委员审核:

1)在已确立的或共同接受教育的背景条件下进行的研究,包括师范教育实践,如关于普通或特殊教育措施的研究,或者关于教学技术、课程、教室管理方法效果的比较研究。

2)关于教育测试的研究(认知、诊断、资质、成就),如果从这些资源中获得的信息是用这样的方式记录的,即客体不能确认,或不能直接从与客体联系的物体中确认。

3)包括调查或面谈的研究,下列除外:

- 反应是用这种方式记录的,即客体能确认,或直接从与客体联系的物体中确认。
- 如果被试在研究之外被广泛知晓,那么他们的作答有可能将他们置于犯罪或民事责任的风险之中,或者对他们的经济地位和就业造成损害。
- 研究涉及被试自己行为的敏感方面,如非法行为、吸毒、性行为或酗酒。

注意,"免除"意味着项目可以由协调者或委员会主席审核,而并非项目可以不需要 ACHE 的审核同意而执行。

快速审核。由联邦规则章程规定的项目的特定方面可以快速审核。PIs 可以在以下三种项目中得到快速审核:

1)由 ACHE 先前同意的项目,而项目只是在协议方面有较少修改。

2)由 ACHE 先前同意的项目,其同意日期已超过了一年。对后一情况,快速审核采取"更新"的方式。

3)一些项目不是可以免除的,但是只是包括很少的风险,如 45CF 46.110(f)(1.7)所描述的。

如果建议符合这些范畴(由联邦规范准则严格定义的),而且申请符合同意标准,同意通知书将立即送给 PI,PI 一接到通知,就应该开始项目研究。

委员会审核

既不符合免除也不符合快速审核范畴的建议将交给委员会的三个成员审核。委员会成员审核申请书及其附件,并将推荐意见及评论交给协调者。这些推荐意见及评论由 ACHE 汇编并整合。然后在全体委员会每月会议中提交审核。在这种会议中,所有提交审核的建议将进行讨论并投票。审核员的审核结果与评论将交给 PI。

委员会活动的形式

ACHE 审核的结果可以采取几种形式。每一种特殊形式讨论如下:

不批准。当委员会认为参与者的风险是不可接受的时候,项目将被否决。风险的可接受性由研究的耗费与效益比决定。如果研究的潜在价值高,研究的高风险也将被容忍。反之亦然。PIs 应当注意只有一小部分的审核项目将被否决。

PI 在摘要部分如果不能充分描述项目,也常常是否决的原因。鉴于此,PI 应当用专业语言描述筛选、安全措施和连贯性的计划,以应对项目程序的不利方面。

暂不批准。在一些情况下,未能提供充分的信息让委员会决定被试遭受的风险的程度。这种情况发生是由于在申请中省略了一些关键的信息。那样,由于申请没有详细描述项目,以至不能提供充分信息让委员会就是否同意或反对作出判断。也有个别时候,不能获得同意的原因是审核成员和预期的被试都不能理解所使用的专业术语。PI 必须提供充分的信息以使建议能在每月的会议上考虑。

有条件同意。ACHE 不同意申请,但要求申请者根据委员会的要求对项目的一些事项进行澄清,或者要求对申请材料的内容或形式进行修改。PI 对其要求作出反应。根据项目意外事故的数目,主席和协调者将作出反应是否充分的决定,或者征求 ACHE 其他成员的意见。如果应答符合委员会设立意外事故原则,项目将会获得同意。

同意。同意常常包括被试风险的评价:①被试不会遭到风险;②被试会遇到一定风险,但研究利益大于风险,根据这种同意的通知书,PI 可以开始研究,部门主席和成员也会获得同意的通知。

更新与已获同意项目的持续审核

请注意:联邦规范限制 ACHE 同意的保有期限是一年。因此,即使已获得同意的

项目也将受全国健康研究所的“持续审核”。在每个项目获得同意满一周年时，PI 将会收到 ACHE 办公室发出的通知，即同意即将到期，并提示提供关于审核的“更新”信息，将 ACHE 的同意延伸一年。

如果研究继续，但 PI 预想同意协议要发生重大变化，项目就必须由全体委员审核。如果项目如计划的一样继续进行，或仅仅与协议有小许改变，那么一个快速审核将认可“更新”并将期限延至下一年。然而，PI 应当注意，全国健康研究所最近强调需要全面的继续审核和同意项目的监督。那样，被全体制度性审核委员会审核的项目在更新同意前将被全体制度性审核委员会成员再次审核。

至少，ACHE 在发布更新同意前需要用一个最近使用的同意形式的副本。最后，如果 PI 报告已经获得同意的项目已经完成，项目将被指明过期或该项目的 ACHE 文件将被归档。

同意形式要求

为了你使用的方便，这儿提供同意形式的例子，其中有一些上标文字，指的是相应的所要求的项目。在应用时，同意形式必须有以下的项目（在你的工作文档中请注意删去上标文字）：

1. 要有研究目的的声明和对接下来研究程序的简洁描述。确认所有在本质上按实验归类的程序，而不必按已经证明或建立的原则归类。即，尚未被证明或确认的。
2. 对被试的任何可预见的不适和风险的描述（生理的、社会的和心理的）。
3. 当有较多风险时，应当包含堪萨斯民事侵权法案的声明。如果包括较多风险，声明应当如下：“在受到伤害的情况下，如果能够证明这种伤害是政府雇员在其职责范围内由于疏忽、错误行为或失职而引起，堪萨斯民事侵权行为方案将提供赔偿。”
4. 从这项研究中，可以合理推断出被调查对象或其他主体利益的情况。
5. 可选择的程序的告知应该对被调查者有利（通常只适用于与药物治疗有关的研究）。
6. 回答调查的询问包括研究主题、联系对象，以及电话号码和地址（关于调查过程、目的或者调查对象的权利的问题）。
7. 声明参与是自愿的，参与可能随时被中断，并且拒绝参与或者决定中断参与将没有惩罚或者损失，这是参与者的权利。
8. 如果有涉保密事宜的话，要声明那些能确认被调查者身份的资料将在何种程度上为被调查者保密。
9. PIs 的姓名，以及他们的部门和电话号码。
10. 除了已同意的被试之外的其他被试的名字（例如那些没有能力作出告知同意的被试）。
11. 被调查者的签名（如果合适）。
12. 如果被调查对象没有满 18 岁，要求征得被调查对象父母或监护人的同意。
13. 如果这种结果没有出现在同意表格，则参考给予被试的任何书面的解释。

14. 在研究中给参与者一个时间上的承诺。
15. 除了被调查群18岁以下的个体外,其他的被调查者应有这样的签名:我签名,确认我至少有18岁。
16. 委员会要求应包括如下内容:我签名,表示我承认我已经得到一份保存的调查表的复印件。这句话可以放在签名线下。
17. 当同意表格的长度超过一页时,页码的形式应包含总的页码数(例如"1/3,2/3,3/3")。
18. 必须包括监督人的名字、部门和部门电话号码。

告知同意声明的例子

注:这两个说明性的例子之前是一个告知同意的要求的目录。在整个这些例子中出现的上标符号对应于出现在上面告知同意的目录列表上的相关要求和要素。

例1 一项与最小的危险有关的项目

堪萨斯大学________系

如下的信息提供给你,以决定你是否愿意参加此项研究。你应该意识到:即使你同意参与研究,也可以随时退出,而不受任何惩罚。

我们的研究兴趣在于:媒体是如何影响人们看待他们自己、他们的问题以及他们的未来的。观看一些录像资料,与研究人员谈话,以及完成一些书面和口头的任务。这些项目将在两个小时内完成。

录像的内容和问题会涉及________,因此你可能对研究中的一些材料和话题感到轻微的不适应。尽管参与不会让你直接受益,但我们相信所获得的信息在评价媒体对观众的影响会有用。[4]

虽然我们强烈希望你的参与,但却是绝对自愿的,[7]我们保证你的名字不会出现在任何研究结论中。[8]信息将只以一个号码的形式鉴定。[8]无论在研究结束前还是结束后,如果你想得到关于研究的其他信息,请随时用电话或E-mail联系我们。[6]

真诚的

John Doc[9]	J. D Smythe, PH. D.[18]
PI	职员监督人
堪萨斯大学人类研究中心	堪萨斯大学人类研究中心
785 864 ________[9]	785 864 ________

____

被试同意参与的签名

我签名,保证我至少已18周岁,而且已持有一份同意表格的副本。

(注:第3、5、10、12项对上述同意表格不适用)

例2　一项不仅只是风险最小的项目

堪萨斯大学________系支持在研究中对人作为被试的保护。提供下列信息让你决定是否愿意参与当前研究。你应当意识到即使你愿意参与,你在任何时候可自由退出,也不会影响你在本系参与其他研究的机会。[7]

该研究还测试你的日常活动的运动能力,测你的平衡能力以及其他能力,研究低强度依靠氧气的运动计划是否能提高你的行动能力和改善你的日常活动。

关节灵活性的测试将使用测量关节角度的工具:角度计,以及"坐并且达到"测试,用来测试脚和背部的灵活性,在你依照我们的指引活动时,通过观察,测试你在四周走动的不同能力,这将需要你的每天一小时的时间,一星期三次,共六个星期。[14]

这种试验调查由堪萨斯大学心理学专业训练有素的研究生进行。所有的程序都经过了________系的顾问的同意。虽然我们强烈希望你的参与,但却是绝对自愿的,[7]在研究完成前,进行中或完成后,请尽管询问与研究有关的任何问题。[6]我们保证你的名字不会出现在任何的研究结论中。[8]名字将会在数据使用前全部删除。

与你的参与相关的预期利益包括你的灵活性和自由行动的能力的信息。当然也有可能(但不是承诺)在参与这个项目后,你在这些方面的能力会有显著提高。

不适与风险是非常小的,在研究开始的时候可能有轻微的半疼痛。由身体活动导致的受伤之类的小事件的可能性很小。[2]然而依据 HEW 规则,我们提供以下信息:

"一旦受伤,且可以证明受伤是由于雇员在其他职责中的疏忽或不作为,堪萨斯民事侵权法案将提供赔偿。"

请依据你对事件的全面了解及事件目的了解,你的预期收益,可能遇到的不适和风险,签署你的同意。我感激你的帮助。

John Doe[9]	J. D Smythe, PH. D. [18]
PI	职员监督人
堪萨斯大学人类研究中心	堪萨斯大学人类研究中心
785 864 ________[9]	785 864 ________

____

被试同意参与的签名

鉴于我已签名,我保证我至少已 18 周岁,而且已持有一份同意表格的副本。

(注:第 5,10,12 项对上述同意表格不适应)

关于告知同意所列条目的任何一项的省略必须明确说明,必须在申请书的文本中标明。根据联邦规则章程(45 CFR 46.116 和 117)中关于条件的例子的讨论中,特定的条款是不适应或不恰当的。

1/98

ACHE ________________(签名)

堪萨斯大学人类实验咨询委员会项目同意申请表

1. 调查者的名字
2. 联合部门
3. a. 学校与家庭邮政地址
   b. E-mail
4. 电话: (学校)________________(家庭)________
5. 项目研究成员的名单:
6. 调查者类型和行为种类(检查合适的分类标准)
   ——堪萨斯州大学的职员或教员
   ——提交计划申请外部基金;代理单位:____________
   堪萨斯州大学/KUCR 项目号码:____________(由 ACHE 完成)

(ACHE 必须把所有已经获得同意申请的协议与相差的 ACHE 申请协议进行比较)

——提交计划申请内部基金;来源:____________
——没有经费的计划
——其他
——堪萨斯州大学学生:________研究生 ________本科生 ________其他
——主题 ____________专题
——班级项目(号码和班级名称):______________
——中立研究(系级管理者姓名):______________
——其他(请解释):
——调查者不是来自劳伦斯校园,但是使用通过堪萨斯州大学而获得的课题

7. a. 调查的名称:________________________
   b. 如果不同于以上的内容,请填写主办项目的名称:________________
8. 除了系、科,或者堪萨斯大学的学生外的个人。请区分调查者和研究群体:

____

9. 授权:

通过 E-mail 等提交了申请,我确信我已经阅读、理解,并遵守堪萨斯大学以人为研究客体的有关政策和程序,我同意标准并将坚持 ACHE 的政策和程序。我熟悉那些已经出版的符合伦理的处理被试的方针,这些被试与我研究的特殊领域有关。

日期:________________ 日期:________________
签名:________________ 签名:________________
第一调查者 系管理者
签名:________________ 签名:________________
第二调查者 第三调查者

首要调查者:________________　　　　　ACHE#:________________

项目名称:____

10. 请回答下列关于研究建议行为方面的问题:(请填写“是”或者“否”)如果回答“是”,一定要在摘要和同意表格中提供详情。

研究是否包括以下方面:

——a. 毒品或者其他受控制的物品?

——b. 因参与而向被试支付报酬?

——c. 通过合作机构而获得被试?

——d. 内部被被试占有或者外部运用于被试的物品?

——e. 机械的或者电子的装置(如电极)应用给被试?

——f. 从被试身上获取液体(如血液)或者生理组织?

——g. 被试经历压力(生理的或者心理的)?

——h. 对被试关心的课题目标与实施步骤的任何方面存在欺骗(误导或者收回信息)?

——i. 是否有被试能被判断出在自由决定同意与否上是理性和自主的(比如,少数民族、发育迟缓的人,或者那些被送进了专门机构的人)?

——j. 是否有任何步骤或行为有可能使被试处于危险之中(心理的、生理的或社会的)?

——k. 有____会见,____调查,____问卷,____录音,____录像带吗?(检查所有的申请材料)

l. 数据收集时间超过一年了吗?

m. 已经填写的同意表格将被利用吗?

注:ACHE 对同意表格中的弃权选项作出最终决定。

11. 研究涉及被试的大致数量:____________

完成这页的下列问题。请不要将内容写到下页去。

12. 课题目标:__________________

13. 描述选中被试的概况(年龄、性别、种族或者其他典型特征)。

____

14. 介绍被试的选取方法(如果被试是从合作机构招募过来的,例如诊所或者其他服务机构等,被试的姓名和其他私人信息(像医疗诊断),没有被试的书面允许不能去获取)。

____

15. 项目的计划步骤概要(必须在这页完成)。

____

(这里凡是对申请书中问题 10a 到 10m 回答了“是”的条款都必须填写细节问题,如:毒品、合作机构、保护措施和计划完成后对磁带的处理、问卷、调查、其他数据,以及

详细汇报欺骗的程序。)

提交一份完整的申请到 ACHE,233 Youngberg Hall。一定要包含同意表格,问卷和其他申请的支持文件。

作为制度性审核委员会申请者培训项目的组成部分,堪萨斯大学已经建立了一个网上指南,它说明了制度性审核委员会申请者请求批准时需要知道的评审原则。这能够在网站 www. research. ukans. edu/tutor 上找到。

另外,制度性审核委员会管理者全国联合会在制度性审核委员会经营中提供经过和证明。联合会可以通过 www. naim. org/或者(404)766-9890 联系。

关于告知同意的推荐读物

Akabayashi, A. Fetters. (2000). Paying for informed consent. *Journal of Medical Ethics*, 26, 212-214.

Chastain, Garvin, & Landrum, Eric R. (Eds.). (1999). *Protecting human subjects: Departmental subject pools and institutional review boards*. Washington, DC: American Psychological Association.

Hecht, Jeffrey B. (1995). *The institutional review board in social science research*. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Chicago.

Kuther, Tara L. (1999). Competency to provide informed consent in older adulthood. *Gerontology and Geriatrics Education*, 20, 15-30.

Miller, Catherine. (1995). Protection of human subjects of research in Canada. *Health Law Review*, 4, 8-16.

Parascandola, Mark, Rusnak, Eileen, Saunders, Carol A., & Sugar, Alan M. (2000). *Foundations of human subject protection*. Wellesley, MA: Center for Clinical Research Practice.

Pinals, D. A., & Appelbaum, P. S. (2000). The history an current status of competence and informed consent in psychiatric research. *The Israel Journal of Psychiatry and Related Sciences*, 37, 82-94.

Sieber, Joan E. (Ed.). (1984). *NIH readings on the protection of human subjects in behavioral and social science esearch*. Westport, CT: Greenwood.

Sieber, Joan E. (1992). *Planning ethically responsible research*. Newbury Park, CA: Sage.

Stephens, John Patrick. (2000). Is this informed consent? *Psychiatric Bulletin*, 24, 154.

White, G. (2000). Informed consent. *The American Journal of Nursing*, 100, 83.

4

QUALITATIVE RESEARCH

定性研究

当研究者对某一现象特别感兴趣，并想获得大量有关这一现象的具体知识时，定性研究方法常常会被运用。在本书这一部分，Creswell & Maietta 把这些方法分为五类：叙事研究（narrative research）、现象学（phenomenology）、扎根理论研究（grounded theory research）、民族志（ethnography）和个案研究（case studies）。他们逐一探讨这些方法的关键内容，以及如何运用这些方法来回答与定性研究最相称的问题。

有一点需要注意，尤其是那些初学者：定性研究方法并非定量方法的替代品，虽然同本书前半部分所讨论的定量研究方法相比，它发展得稍微迟了一点。定量研究方法与定性研究方法不是回答同一问题的不同方法，相反，这两种方法构造不同的途径来回答不同类型的问题。在理解如何最好地研究个体和群体行为时，它们有各自不同的世界观以及由这些世界观派生的潜在假设。

导 论 4.1

人们通常会认为定性研究就是灵活而开放地理解被研究对象所持的观点和意义。多少年来,无论是研究者申请研究经费,研究生同导师讨论问题,还是把成果用于发表,只要是使用定性方法,就意味着会被斥为在呈述他们的研究时缺乏系统的程序。随着定性研究“后现代”转向——这一流派强调研究者的自我意识,降低对方法的关注,生产相对性的知识——的到来(Denzin & Lincoln),定性研究才得到进一步的发展。

但有人会认为定性研究者不使用可供检验的程序来收集和分析数据。得承认,有些定性方法在程序方面的确不够结构化和系统化,但有些用于定性分析的程序已经出现,尤其是当研究者分析大规模的数据库,或者用计算机程序来进行数据分析时。

本书这一部分主要介绍定性研究的方法,聚焦于五种定性研究途径。这五种途径是:叙事研究、现象学、扎根理论、民族志和个案研究。在讨论每一种途径时,我们将简单地回溯一下它的起源,对它进行定义,探讨其变量,指出其关键的要素,并展示其主要的研究程序。

系统的调查方法 4.2

对于那些进行定性研究的人来说,调查方法确实多到了难以选择的地步。稍微浏览一下几种类型学,我们可以认识到这一点。一种较为流行的分类是由 Tesch (1990)提出的。依据调查者的主要兴趣,他将 18 种方法归入 4 大类。Wolcott (1992)用一个树形图来标示 25 种收集数据的方法。根据研究者所关注的人类生活的领域,如个体、社会或文化,Miller 和 Crabtree (1992)建构了 18 种类型。在教育领域,Jacob 将所有的定性研究按传统分门别类,如生态心理学、符号互动论、整体民族志。Lancy (1993)用学科的视角来将定性研究的方法归类,如人类学、社会学、生物学、认知心理学和历史学。Denzin & Lincoln (2000)区分了几种研究者用来收集和分析实证材料调查的策略。他们提出了表演民族志、个案研究、民族志、诠释实践、扎根理论、生活史、新历史、见证法、参与行动研究以及临床模式等作为重要的研究策略。

现在定性研究者有太多的调查方法可以选择。我们将会选择性地讨论五种方法。最近的研究显示,这五种方法已具备了较系统的研究程序。虽然还有一些研究方法正在出现和发展,但它们或多或少缺乏我们讨论所需要的清晰的程序。在挑选这五种方法中,我们还认识到这些方法在社会科学研究中是经常应用的,并经常出现在社会科

学和人文科学的学术期刊中。

更重要的是,我们的选择是基于这样一种兴趣,即从不同的社会科学传统来确定研究方法,以便于鼓励定性研究者超越他们自己的学科,观察那些其他领域中的途径。为达到这一目的我们从人文学科中选择了叙事研究,从心理学和哲学中选择了现象学,从社会学中选择了扎根理论,从社会学和人类学中选择了民族志,从人文、社会科学以及诸如评估研究的应用领域中选择了个案研究。

4.3 关于五种方法的理论观点

从每种传统的简略描述中,我们可以看出各种类型的定性研究之间的基本的区别。表4.1列出了几个维度。在最基本的层面上,这些方法在研究焦点或者说研究的最初目标方面各有不同。探索一个生命的故事同产生一个理论或者描述一个文化群体的行为是截然不同的。而且,尽管在学科的起源上存在重叠交叉,但是一些方法较少被用于其他学科(如扎根理论源于社会学,民族志源于人类学和社会学),而另外一些则广泛用于多个学科(如叙事研究和个案研究)。收集数据的过程也有所不同,有些侧重于运用某些特殊的数据收集方式(如民族志更多地使用观察法,访谈法则更多地被用于扎根理论),有些则综合运用各种方式(例如,个案研究往往用多种方式收集数据来呈现一个深度个案)。这些差异尤其在数据分析阶段表现地更为明显。不仅在分析的严格性方面存在差异(如扎根理论最为严格,叙事分析则相对松散),而且分析步骤的复杂性也有差异(例如,现象学的步骤很多而民族志的步骤很少)。

其结果是,每一种传统,包括它的叙述形式,都受到它先前发展过程的影响。详细地描述个体故事是叙事研究的基础,而描绘某类现象及其经验之本质则成为现象学。扎根理论往往通过描述一个实质性的模型来构造理论;民族志整体性地描述一个社会文化群体或系统;而个案研究则对一个有边界的系统或个案进行深度研究。有了这样一些简单的了解,我们现在转向对叙事研究的探讨。

表 4.1 定性研究的五种途径

维　度	叙事研究	现象学	扎根理论	民族志	个案研究
焦点	收集有关过往经验的故事	理解围绕某一现象之体验的本质	从田野资料中发展出理论	描述或理解文化或社会群体	对一个或多个个案进行深度分析
学科起源	文学 历史 心理学 社会学 人类学	哲学 社会学 心理学	社会学	文化人类学 社会学	政治科学 社会学 评估研究 城市研究 其他社会科学
资料收集	主要使用访谈法和文献法	对 10 个左右的人进行长时间的访谈	访谈 20 ~ 30 个人使种类"饱和"并细化理论	在长时间(比如半年到一年)的田野工作中使用观察法、访谈法和实物收集来收集资料	利用多元的资料,如文献、档案记录、访谈、观察、实物收集和定量数据
资料分析	讲述故事 重构故事 指出主题 描述背景	陈述语句 分析意义 概括描述体验	开放式编码 轴式编码 选择性编码	描述 主题分析 解释	描述 分析主题 呈述命题
叙述形式	按时间顺序讲述一个人的生活故事	描述体验的"本质"	构造理论或理论模型	描述个体或群体的文化行为	深度研究一个或多个个案

叙事研究 4.4

尽管叙事研究受到颇多的关注,也经常在文献中被广泛讨论,但是作为定性研究之一种,其方法和程序仍在发展之中(Errante, 2000)。在叙事研究中,研究者描述个体的生活,收集并讲述故事,并且记录个体体验的叙述(Connelly & Clandinin, 1990)。作为定性研究的一种独特的形式,叙事研究尤其强调对个体的研究:收集有关个体的故事,记录个体的体验,并分析这些体验对当事人的意义。

叙事研究被广泛用于社会科学。Riessman(1993)所谓的"叙事的转向"包括了所有的人文学科,因此这一研究形式不是某一特定研究领域所独有的。很多学科,包括文学、历史、人类学、社会学、社会语言学以及教育学,都有人声称在搞叙事研究,并且已经发展出若干与学科相关的具体路径。例如近来被社会科学讨论的肖像画法的艺术与科学(Lawrence-Lightfoot & Davis, 1997)。这种研究方法要求列出个体的特征,记录他们在具体社会文化背景下的声音和形象。又例如在教育研究领域,Clandinin &

Connelly (2000)做了第一个有关叙事研究用于教育学的研究综述。在《经历的故事和叙事研究》(*Stories of Experience and Narrative Inquiry*,Clandinin & Connelly,1990)这篇很经典、涉及面极广的文章中,Clandinin & Connelly 引用了很多社会科学运用叙事方法的研究,详细阐述了收集叙事田野笔记的过程,并讨论了叙事研究的结构和写作方法。在课堂教学和讨论的基础上,这篇文章拓展了他们先前对叙事研究的讨论(Clandinin & Connelly,1988)。最近,这两个人把他们的最新想法收录在《叙事调查》(*Narrative Inquiry*,Clandinin & Connelly, 2000)中,公开宣称要回答"叙事研究者做什么"(p.48)。

叙事研究之所以如此流行要归功于若干因素。Cortazzi (1993)认为人有兴趣进行个人反省,了解人们之所知,并鼓励人们讲出他们的故事。例如,已婚女性乐于讲故事给孩子、年轻姑娘以及圈子里其他女人听。越来越多跨学科的社会科学研究者极有兴趣从事叙事研究,他们提供了若干程序性的指导,以使叙事研究成为定性研究之一种(详见:Lieblich, Tuval-Mashiach, and Zilber, 1998;Cortazzi 1993;以及 Riessman, 1993)。从 1993 年开始,"生活的叙事研究"(Josselon & Lieblich, 1993)系列丛书每年出版一册,这极大地鼓励了叙事研究在跨学科研究中的努力。

4.4.1 不同的取向

因为很多不同背景的研究者都进行叙事研究,所以叙事研究中有很多不同的取向是可以理解的。目前,叙事研究可以被视为各种不同研究实践的集合体(Casy, 1995/1996),每一种类型都有些许差异。在这里我们将讨论以下四种类型,它们是:传记、生活史、教师—学生研究法,以及理论取向的叙事研究(一个更详细的关于叙述研究形式的讨论,见:Creswell, 2002)。

(1)传　记

在叙事研究中,由谁操刀记录个体的故事关系重大。作为叙事研究之一种,传记由研究者来记录传主的生活经历。传记主要利用档案资料(Angrosino, 1989),不过研究者也可以利用其他形式的资料,如访谈和照片。在自传中,叙述和写作都是由传主自己完成的。自传在教育研究中很少见,但最近有些教师的叙述被报道出来(Connelly & Clandinin, 1990)。

(2)生活史

在人类学中,生活史完整描述个体的一生。一则生活史就是对一个人全部生活经历的叙事研究。从事生活史研究的人类学家要去了解在某一文化背景下个人的生活。他们尤其关注个体生活中的转折点或其他对个人生活尤其有意义的事件(Creswell, 2002)。相对而言,叙事研究通常不要求记录完整的生活,而是关注某个生活片段或单个事件。生活史则是通过对单个或多个生活片段、具体情景以及周遭文化的研究,来讲述个人的经历和故事(Denzin,1989)。Clandinin & Connelly(2000)将个人的体验拓展至社会层面,认为体验的本质既是个人的也是社会的,学校的老师和教学也不例外。

（3）老师和学生叙事

另一种类型的叙事研究关注由谁来提供故事。这一方法与教育研究关系尤其密切，在此类研究中，教育工作者是叙事研究的焦点（Creswell, 2002）。例如，教师讲述的是他们个人的有关他们在课堂中体验的故事，作为教育学中叙事研究之一种，教师的讲述被广泛用于概括教师的生活以及用来检验课堂的教学（Connelly & Clandinin, 1988，1999）。另一种叙事研究关注课堂中的学生，在学童的讲述中，学生被要求口头或书面呈现有关他们自己学习的故事（Ollerenshaw & Creswell, 2000）。在教育研究中，很多人可以提供故事，例如管理层、学校董事会成员、门卫和食堂人员，以及其他教职工。

（4）理论取向的叙事研究

最后一个影响叙事研究特征的选择是，是否以及在何种程度上研究者可以运用理论的视角来进行叙事研究。理论视角意味着研究在某一指导性理论或意识形态的框架下展开。在这类研究中，研究者倾向为某一群体或个体辩护。比如，女性主义者从女性的角度讲述女性自己的故事，或者研究者收集边缘群体的故事，为弱势群体呐喊。在所有这些例子中，叙事研究者为那些无法在研究文献中发出自己声音的群体提供了一个发声的机会。

4.4.2 叙事研究的关键要素

尽管叙事研究有很多不同的形式，它们仍有共同的地方。从事叙事研究的人强调从所有研究参与者中学习并获得知识。当一个或若干人开始讲述故事的时候，这种学习就出现了。对 Clandinin 和 Connelly 而言，叙事研究中的故事同时反应了个人的体验和社会的体验。对体验的强调源于杜威的哲学思想，他认为个人的体验是理解一个人的关键。杜威认为体验是连续的，一个接着一个的（Clandinin & Connelly, 2000）。这些故事构成数据，研究者可以通过访谈和非正式的谈话获得。这些故事，或曰田野文本，就是原始资料。研究者根据原始资料中的问题、角色、情景、行动以及结果来重新讲述或构造故事（Ollerenshaw & Connelly, 2000）。重构故事是一个过程，需要先收集故事，分析故事中的关键要素，然后根据时间顺序重写这些故事。通常情况下，被访者在讲故事的时候，这些顺序是缺失的或不合逻辑的。通过重构故事，研究者为这些故事建立了因果关系。在这一过程中，叙事研究者会加进去很多有关故事的背景和具体情景，比如被研究者的朋友、家庭、工作地点、房屋、社会组织或学校。总之，交待故事所发生的具体环境。

叙事研究中的故事都是以第一人称的方式讲述或重述个人或社会的体验以及相关事件。通常这些故事有开头、过程及结尾。像好的小说一样，这些故事也要具备以下这些基本要素：困境、冲突或斗争，主角或人物，以及如何进行筹划来突破困境的因果顺序（Carter, 1993）。在更广义的层面上，这些故事应该如同小说一样，包含时间、地点、事件和场景（Connelly & Clandinin, 1990）。研究者在讲述故事的时候通常要指出其中的主题。这样，分析定性数据同时包括描述故事和讲述其中的主题。另外，研究者经常按时间顺序来重新组织故事，并把个人过去、现在和未来的经历置于一个具

体的情景和背景中来描述。Cortazzi 认为正是对时间顺序的强调使叙事研究有别于其他研究方法。

在收集和分析数据的过程中,研究者与参与者需要一起合作,检验故事并讨论其中的意义。在被研究者的故事中,可以交织着研究者的洞见。这种合作在研究过程中包含很多层面,如共同确定有待检验的主要现象和田野文本的形式,乃至共同决定研究者如何最后呈现重构出来的个体经历。合作不仅意味着协调研究者和参与者之间的关系以便缩小参与者所述与研究者所呈之间的差异,还意味着解释研究对于参与者的意义,讨论从资料收集到写作的转变,以及确定参与者涉入研究的方式(Clandinin & Connelly, 2000)。

4.4.3 叙事研究的程序

以下是叙事研究的一些较重要程序,从中我们可以窥见叙事研究的过程。

1)确定一个研究问题,重点探讨一个或多个人的生活经历,以及由此引出的个人或社会的故事。研究的目的在于提供一个翔实的故事。

2)选择一个或多个有故事可讲的人,花一定的时间访谈和记录他们的故事,或者让参与者用日记或其他笔记的方式记录下自己的故事,研究者也可以直接观察他们的生活并做田野笔记。还有一些其他的收集资料的方式,比如:收集研究参与者寄来的信;从参与者的家庭成员那里了解他们的故事;收集有关参与者的材料,如备忘录或其他正式的信件;或者获取相片、纪念品和其他记录个体的生活体验的形式,如舞蹈、戏剧、音乐、艺术和文学(Clandinin & Connelly, 2000)。

3)收集完参与者的故事之后,需要按照时间顺序来重新讲述这些故事。按时间顺序意味着需要呈现故事之主题的过去、现在和未来,所有这些都需要在翔实的背景资料和具体情景中讨论。更重要的是,除了讲故事,还要详细讨论故事所反映的主题及其意义。

4)与参与者合作,让他们积极参与到研究中来。他们可能会帮助你修改研究问题,帮你找资料来支持他们的故事,同你一起分析资料,并协助你完成最后的研究报告。

4.5 现象学

叙事研究主要关注一个人的生活体验,而现象学围绕一个概念或现象来描述多个人的生活经历及其意义。作为探寻人类体验中之意识结构的一门学问,现象学根植于胡塞尔的哲学并光大于 Heidegger, Sartre 和 Merlear-Ponty 的哲学论述(Spiegelberg, 1982)。现在,现象学已经被广泛用于社会和人文研究,尤其是社会学(Borgatta & Borgatta, 1992; Swingewood, 1991)、心理学(Giorgi, 1985; Polkinghorne, 1989, 1994)、护理与健康科学(Nieswiadomy, 1993; Oiler, 1986)和教育学(Tesch, 1988)。

现象学起源于德国哲学家胡塞尔。从1913年开始直到他退休,胡塞尔写了很多著作来阐述现象学哲学。由于胡塞尔的思想非常抽象,直到1945年, Merleau-Ponty在《理解的现象学》(*Phenomenology of Perception*)中还在问:什么是现象学?众所周知,胡塞尔把所有东西都冠之以"现象学"(Natanson, 1973)。

胡塞尔强调很多要点(Moustakas, 1994; Natanson, 1973)。研究者需要探寻体验之本质性的不变的结构,并强调意识的意向性;在意识的意向性中,体验既包括外在的形式又包括内在的意识,如记忆、映像和意义。现象学资料分析超越了化约主义之方法论,转而分析具体的陈述和主题,并寻找所有可能的意义。研究者也要避免所有的偏见来概括他们的体验,需要依靠直觉、想象和普适性结构来得到一幅有关体验的图画。

从这些哲学论述中,我们可以辨识出以下四点(Stewart & Mickunas, 1990):

1)回归哲学的传统任务。在19世纪末,哲学已经局限于用经验手段来研究这个世界,即科学主义。回归哲学的传统任务就是,不再被经验科学所局限,回到古希腊哲学对智慧的追求上去。

2)没有预设的哲学。现象学的研究路径要求对所有何谓真的判断存而不论,直到找到一个更为确定的基础。胡塞尔把这种对预设的悬置称为"epoche"。

3)意识的意向性。这一主张表明意识总是被指向某一对象。于是对象的真实性无可避免地与一个人的意识纠缠在一起。胡塞尔认为,事实不能被分割为主体或客体,如此一来便超越了笛卡尔式的二元对立,转而关注浮现于意识之中的对象的意义。

4)拒绝主体、客体的二分法。这一主张是意识之意向性的自然延伸。只有将对象置于个体体验之意义的背景中才能理解其真实性。

4.5.1 不同的取向

在社会科学各个学科中,尤其是社会学和心理学,很多人赞成以上的哲学主张,并把它们付诸于研究。于是,各种不同于哲学现象学的形式发展了出来,如反思现象学、对话现象学、经验现象学、诠释现象学和社会现象学(Barritt, 1986; Tesch, 1990)。我们将简要介绍一下社会现象学,然后主要讨论心理现象学。

社会学取向的现象学主要归功于舒茨(Schutz)。他对社会事实的研究清楚地表明了现象学的本质(Swingewood, 1991)。舒茨对社会的普通成员如何构造日常活动的世界很感兴趣,尤其是个体如何有意识地赋予社会互动的意义。作为舒茨思想的延伸,加芬克尔(Garfinkel)把这种方法称为民族方法学(ethnomethodology)。民族方法学主要借用民族志和文化研究的方法,分析日常对话,并探讨个体如何产生日常生活的意义(Swingewood, 1991)。

心理学取向的现象学同样关注体验的意义,但它关注的中心点是个体的体验而不是群体的体验。如《现象学中的迪凯纳研究》(*Duquesne Studies in Phenomenology*)一书所述,心理现象学的核心是:"确定体验之于个体的意义,此个体曾经经历并有能力描述这一体验。从个人化的描述中找出具有普适性的意义,或者说此一体验的结构本质。"(Moustakas, 1994, p.13)

Moustakas (1994)发展出反思现象学,这一方法可以追溯到胡塞尔,但它更强调放弃研究者的预设,并从体验之内容与方式中探寻具有普适性的结构。

4.5.2 现象学研究中的主要要素

很多人已经讨论过如何操作心理学现象学研究,如 Duke (1984), Tesch (1990), Giorgi (1985, 1994), Polkinghorne (1989),以及 Moustakas (1994)。大家在这个问题上基本上有了一些共识。不过这些方法仅仅"把这些现象学原则作为研究指南或大纲,而研究者在研究的时候总是希望发展出特别适于理解特定现象的具体研究计划。"(Polkinghorne 1989, p. 44)有了这样一些提醒,我们可以来概述一下现象学研究程序的主要组成部分。

研究者需要理解这一方法背后的哲学意义,尤其是关于在研究人们如何体验某一现象时。Epoche 就是中立,要求研究者放弃专家词汇,从被研究者的角度去理解有关现象(Field & Morse 1985)。在探讨体验之于个体的意义时,研究者设计问题并要求被访者根据问题描述自己的生活经历。然后研究者从这些回答中收集资料。通常,此类数据主要通过很长的访谈来获得,被访者 5 ~ 25 个人不等,研究者还可以加上自己的反思和先前已有的研究结果。资料会被进一步分析,发展出一些较普遍的主题,然后形成一个详细的有关被研究现象的报告。如果读过这些报告的人认为"能更好地理解那些有此类经历的人",那么就达到目的了(Polkinghorne, 1989, p. 46)。

4.5.3 现象学研究的操作程序

以上这些要素对现象学研究程序很重要。在步骤方面,现象学分析大体上类似于心理现象学家所讨论的内容(Moustakas, 1994; Polkinghorne, 1989).

1)确定一个主要的现象(如悲痛的意义)来研究并探讨其本质,或经历此现象的人赋予它的普遍性意义。

2)提出有关此意义的中心问题。例如一个现象学家会问:"个体赋予悲痛以什么样的意义? 对于他们来说,悲痛是什么? 在何种程度上他们体验到悲痛?"

3)主要通过访谈收集数据,尽管其他的方式,如观察法以及对文献、物品、音乐和诗歌的分析也会被用于现象学研究(关于多种方式的数据收集,参看:Lauterbach, 1993)。

4)Moustakas (1994)确定了分析数据的步骤。先是确定"有意义的陈述",用被访者所用之词汇(或研究者自己的语言)来概括所要研究之现象的意义。

5)把众多有意义的陈述简化为意义单位或主题。在这一过程中,研究者先寻找重复或多余的意义陈述,然后把他们合并起来成为意义单位或主题,并用这些单位或主题来描述现象之于个体的意义。

6)分析个体体验这些意义单位或主题时的背景。对现象学来说,重要的不仅仅是找出个体体验的内容,更要搞清楚其背景和当时的情景。

7)反思研究者自己对相关现象的个人化体验。有些现象学家采用这一步骤。包括主动记录下并反思你自己的体验。这些文字信息通常涉及研究者如何经历哪一现

象以及在哪种背景下经历。

8)详细分析参与者之体验的本质,并写出来。从分析意义陈述,意义单位,到分析体验的内容(文本描述),到呈现体验的方式(结构陈述),研究者写出一个详细的有关参与者体验之本质的文本。有时,参与者的陈述会根据一些具体特征单独分析,如男性和女性(Riemen, 1986)。对于现象学家来说,在研究的结尾就应该指出参与者体验的"本质",以及体验的具体环境。这部分也可以包括研究者自己有关这一现象的体验。

扎根理论研究 4.6

现象学研究强调体验之于某些人的普遍意义。与此相对照,扎根理论研究意欲产生或发现一个理论来解释某一概念、过程或人际间的互动。在扎根理论中,研究者主要收集访谈资料,多次进行田野访问,寻找不同类型的信息并找出其间的联系,然后产生理论性命题或假设,或者直接呈现一个理论。

Barney Glaser 和 Anselm Strauss 这两个社会学家最早在 1967 年阐述了扎根理论,并在随后的一系列著作中详细论述了这一研究方法(Glaser, 1978; Glaser & Strauss, 1967; Strauss, 1987; Strauss & Corbin, 1990, 1998)。与社会学强调理论导向的优先性不同,他们认为理论应该是"扎根"于田野数据,尤其是人们的行动、互动及其过程。尽管这两人曾经合作过一些著作如《濒死知觉》(*Awareness of Dying*)、《濒死时刻》(*Time for Dying*),但现在他们已经分道扬镳。Glaser(1992)甚至批评了 Strauss 有关扎根理论的取向。尽管存在差异,这两人还是各自出书按照各自的取向来阐述扎根理论。扎根理论也在社会学、护理学、教育学以及其他社会科学领域流行开来。最近,Charmaz (2000)提出扎根理论的"建构主义"取向,这样,在程序的讨论中又多了一个视角。

4.6.1 不同的取向

从以上简要的历史介绍中我们可以了解到,在涉及如何进行扎根理论研究时有三种不同的取向:Strauss & Corbin (1990, 1998)主张的系统程序取向,Glaser 主张的浮现设计取向(the emerging design)以及 Charmaz (1990, 2000)的建构主义取向。

(1)系统设计

扎根理论中的系统设计取向被广泛运用于社会科学研究。Strauss & Corbin (1990)提出了严格而具体的程序,并在第 2 版中详细阐述了构造扎根理论的技术和程序(Strauss & Corbin 1998)。在程序指导方面,这本书比最初介绍扎根理论的那本(Glaser & Strauss 1967)更具有指导意义。扎根理论的系统设计取向重视采用开放的、轴式的以及选择性的编码,在此基础上进行数据分析,然后发展出一个逻辑范式或描绘出一个理论的图像。

在开放式编码时,研究者先得从资料中分辨出若干范畴或主题。在每个范畴内,又得找出若干可被视为亚范畴的"特质"。这些"特质"在扎根理论中被维度化,以尽可能地将这些特质在一连续体中表示出来。例如,学术委员会或系主任,作为一个供给者(范畴),为全体教员提供资助(特质),实际的资助就构成一个可能的连续体,从大型的启动经费到无足轻重的交通补助(维度化的特质)。

在轴式编码时,扎根理论家选择一个开放的编码范畴,把它置于被研究之过程(或互动)的核心,然后把它与其他范畴相联系。这些类型包括因果条件(影响核心现象的因素)、策略(回应核心现象所采取的行动)、背景及影响因素(广义和狭义上影响策略的情景因素)和结果(使用这些策略的成果)。这一阶段往往需要画一个图,也叫做编码范式,来把各种范畴联系起来。

在选择性编码中,扎根理论家需要构造出一个理论,来解释各范畴之间的相互关系,也为所研究的过程或互动提供一种抽象解释。扎根理论家写出连接不同范畴的故事线索,然后根据个人对有关理论的了解来拣选理论。从程序上说,这是一个整合和重塑理论的过程(Strauss & Corbin, 1998)。在构造故事线索时,研究者可以检验哪些因素影响到现象?如何影响?如何导致对某些特殊策略的使用并产生某些特定结果?

这三种编码程序的运用意味着系统扎根理论家用固定的程序来构造理论,依靠分析轴式编码中的特定类型的范畴,以及使用图的形式来呈现他们的理论(Creswell, 2002)。这一取向的扎根理论可能会以假设(Strauss 和 Corbin 称之为命题, 1998)来结束研究,并把轴式编码中各范畴之间的关系清晰化。

(2)浮现式设计

尽管 Glaser 与 Strauss 合作出版了最早有关扎根理论的著作(Glaser & Strauss, 1967),Glaser 对 Strauss 现在的研究取向持批评态度,并写了一个长篇批评文章(Glaser, 1992)。在这篇文章中,Glaser 认为 Strauss 和他的合作者 Corbin 过分强调规则和程序,有关范畴的预想性框架,以及证明理论而不是生产理论(Babchuk 1996, 1997, 很好地综述了扎根理论在过去如何并将继续在社会科学研究中运用)。另一方面,Glaser 强调让一个理论从资料中浮现出来,而不是使用特殊的、预定的范畴,例如轴式编码图所展示的那样(如,因果条件、内容、影响因素、策略和结果)。而且,在 Glaser 看来,扎根理论的目的就在于让一个研究者去解释"基本的社会过程"。这种解释需要不断地进行比较性编码,比较事件与事件,事件与范畴,范畴与范畴。研究的重点在于寻找范畴之间的联系并浮现出理论,而不是简单地描述范畴。最后,研究者构造一个理论并讨论范畴之间的关系,但并不需要画一个具体的图或表。

Glaser 提出了另一种形式的扎根理论,它更加通融与灵活,但相对而言比较缺少指导性。这种形式的扎根理论可以概括如下:

1)扎根理论存在于最抽象的概念层次而非最低的抽象层次,如生活史和简单资料呈现中可找到的那样。

2)理论扎根于资料但并不一定要范畴化。

3)一个好的扎根理论必须满足四个主要标准:合适、可用、中肯,以及可改进。一个根植于坚实领域的细致理论在参与者、实践者和研究者看来应该是与事实相符才

对。如果一个扎根理论可用,那么它应该能够解释参与者的行为及其变化。如果一个理论合适并可用,那它一定是中肯的。最后,理论并非一成不变,当新的数据出现时,理论也会随之修改。

(3)建构主义取向

Kathy Charmaz (1990, 2000)是建构主义取向的主要拥护者,其哲学姿态介于实证主义(更量化一些,如 Glaser, Strauss 和 Corbin 的研究)与后现代研究(质疑方法的重要性)之间。聚焦于研究中参与者所赋予的主观意义,Charmaz 对个体的观点、价值、信仰、感情、设想和意识形态更感兴趣,而不是急于收集事实和描述行动。她认为,如果扎根理论总是使用一些复杂的词汇或术语,图表或者概念示意图来研究体验的话,那将使体验本身晦暗不明,而且这种做法本身就是试图获得话语权力。她主张使用积极的编码,如"重铸生活"。另外,扎根理论程序也不应该把研究者在研究过程中的角色最小化(Charmaz, 1990)。研究者把相关的问题以及相应的"社会学概念储备"(Charmaz, 1990:1165)引入数据分析,研究者也同样把价值、体验和优先性带进研究。此外,Charmaz 还认为,任何由扎根理论得出的结论都是提示性的、不完整的和不确定的。

4.6.2 扎根理论的关键要素

扎根理论家试图广泛地揭示个体之间的行动、互动和社会过程。为了达到这一目标,他们不是依靠现有的理论,而是生产或发展他们自己的主题或范畴。尽管多种方法可以被用来发展这些理论,扎根理论家更多地使用访谈法,让被访者详细讲述他们的体验。研究者根据理论的相关性选择受访个案,这意味着这些人能够为未来的扎根理论提供丰富且直接的洞见,因为他们有相关的体验。典型的做法是,研究者多次进入田野收集访谈数据,通常是访谈 20 ~ 30 个人,得到饱和的范畴(意即进一步的访谈不可能再获得更多的信息)。这些范畴代表信息单位,它们构成活动、事件和实例。研究者也通过观察法和文献法来收集和分析资料,但这些资料形式不是非有不可。

研究者收集资料之始,亦是分析资料之时。实际上,扎根理论的资料收集过程就是一个迂回前进的过程:进入田野收集资料,分析这些信息,重回田野收集更多的资料,再分析这些信息,如此循环往复。在理论导向下选择被访者,这些被访者反过来又帮助研究者更好地生产理论。至于研究者应该几次进入田野,则要看有关信息的范畴是否达到饱和,以及理论是否能详细阐明事实的复杂性。从资料收集中提取信息并从浮现的范畴中比较信息,这一过程被称为数据分析的持续比较法。

按照这一方法收集和分析数据的最终产品就是构造一个理论,一个实质性的关于某一特定问题或涉及特定人群的理论。扎根理论研究的核心就是要发展或生产一个理论来解释所研究的现象及其背景。例如,Strauss 和 Corbin (1990)认为理论就是概念之间的可能关系。扎根理论研究需要在研究的结尾清晰阐述理论,呈现叙事的形式(Strauss & Corbin 1990),一幅具体的图画(Morrow & Smith, 1995),或若干假设或命题(Creswell & Brown, 1992)。在整个过程中,撰写备忘录扮演了重要的角色:研究者通过备忘录记录下正在浮现的理论,备忘录的写作也有助于利用资料修正理论。

4.6.3 扎根理论研究的程序

以下所要讨论的程序主要来自于 Strauss & Corbin (1990, 1998)的著作,他们的研究取向在扎根理论中最为系统,并提供了一个严格且系统的程序性指导。

1)运用扎根理论生产一个宽泛的解释。如果你觉得现有的关于某一现象的理论还不够充分,因为这些理论或者没能抓住现实的复杂性或者忽略了你想研究的人群,那么扎根理论这一研究方法会对你很有帮助。当研究者想发展或修正某一理论,解释一个过程或抽象出人们的行动和互动时,扎根理论尤其受到研究者的青睐。同样地,扎根理论提供一个对社会现象的宏观分析而不是具体的微观描述。

2)考察个体之间的行动、互动或过程。因为扎根理论研究的核心就是解释某一过程,所以你需要首先确定一个你感兴趣的实验性过程。作为研究问题的自然延伸,这个过程可能会在你的研究中出现或变化,但在此阶段你需要对它有个基本的判断,然后去访谈那些正身在其中的人,了解他们的行动和互动的情况。

3)聚焦于一个核心问题。这一问题被用于详细说明某一过程或互动。例如,关心市场营销的扎根理论家会问:"观光者会如何看待历史观光地?"关心教育的扎根理论家会问:"哪一过程能解释学术变迁为什么会在大学和学院发生?"在这两个例子中,研究者都试图解释某一过程。

4)通过访谈来收集数据,被访者要求都经历过你所要研究的过程,因而会对发展理论能有所帮助。虽然扎根理论家可以用多种方法来收集数据,但访谈法最常被用来把握被访者的体验。扎根理论研究的一个重要特征就是研究者多次收集资料,不停地回到田野收集更多的资料,直到所需要的范畴达到饱和,理论被完整地构造出来。收集资料的过程并没有严格的时间限制,研究者需要自行决定何时可以结束资料收集。对研究生来说,比较普遍的做法是访谈 20 ~ 30 个人(Creswell, 1998)。当然,这一规则是灵活的,如果研究者采用多种方法,如观察法和文献法的话,那么就没必要访谈这么多人。

5)采用 Strauss 和 Corbin (1990, 1998)所倡导的三步骤法来进行系统编码。Strauss 和 Corbin 认为,分析资料的过程包括开放式编码、轴式编码和选择性编码。数据编码的过程应该在数据收集阶段就应该开始,以便研究者决定下一步要收集哪些资料。通常数据分析从辨别开放式编码范畴开始,然后用比较法对比数据之于事件,事件之于范畴,直到范畴达到饱和。通常找到 10 个范畴就可以了,当然具体的范畴数目取决于数据库的大小以及所研究对象的复杂程度。

从开放式编码开始,研究会逐渐过渡到轴式编码,并制作一个编码图。这一过程包括从开放式编码中选择一个核心范畴,并把它置于编码图中心。到了这一步,研究者可能重回田野去收集数据,并找出若干与这一核心范畴相关的范畴:原因范畴(什么导致核心范畴),背景范畴(核心范畴是在哪种背景下出现的,受到哪些情景影响),策略范畴(人们如何应对),以及结果(人们策略性行动的后果如何)。以上这些可以用编码图的形式或形象的图示法展示出来。在绘图的时候,研究者可以用箭头的方式标出这一过程发生的顺序。

编码的最后阶段是选择性编码。研究者在这一阶段开始构造自己的理论。这需要把编码图中的范畴相互联系,重新诠释轴式编码图,呈现出一个模型或理论。研究者也可能提出几个命题供进一步研究检验,当然是否会有人检验这些命题或把他们用于实践则是另一回事,在这一点上也存在争议。总之,扎根理论可能会以一系列命题或亚命题的方式呈现出来,讲述一个有关范畴之间相互关系的故事。

6)以呈现命题或陈述范畴之间关系的方式结束扎根理论研究。扎根理论研究往往以命题或具体假设的方式来结束研究,如原因范畴是如何影响核心范畴的?而核心范畴又是如何反过来影响策略范畴的?这些假设可以用比较正式的指导性假设形式表达,如"教员的情绪越低落,越多的系主任会扮演支援性角色并为资浅教员寻找指导者";也可以用讨论的方式表达,陈述模型中各范畴之间的关系。

民族志 4.7

虽然扎根理论从对多个人的研究中发展理论,但这些人并不共享相同的行为模式。对一群有共享文化模式的人进行研究称为民族志,它描述和解释一个文化或社会群体,研究这一群体可被观察和学习的行为模式、习俗和生活方式(Harris, 1968)。民族志既是研究过程又是研究结果,大量出现在学术期刊和著作中。作为一个研究过程,民族志要求对所研究群体的长期观察,通常是参与观察:研究者融入被研究者的日常生活,或进行一对一的访谈,了解他们日常行为的意义、语言以及他们与其他文化群体的互动。

民族志起源于文化人类学。20 世纪初的那些文化人类学家如 Boas、Malinnowski、Radcliffe-Brown、M. Mead 曾用民族志来进行跨文化研究。尽管他们也视自然科学的模式为圭臬,但他们的研究不同于传统的坐在扶手椅上的研究,而是进入所谓的"原始"部落去收集第一手的资料(Atkinson & Hammersley, 1994)。在 20 世纪的 20 年代和 30 年代,一些芝加哥大学的社会学家如 Park、Dewey 和 G. Mead 用文化人类学家的方法研究了美国的一些文化群落(Bogdan & Biklen, 1992)。最近,很多不同的取向在民族志中涌现,例如结构功能主义、符号互动论、文化与认知人类学、女性主义、马克思主义、民族方法学、批判理论、文化研究和后现代主义。这些取向带着各自不同的理论目的来描述和解释文化和社会群体,因而也导致民族志缺乏一个正统。

4.7.1 不同的取向

民族志有很多种,如忏悔民族志(confessional ethnography)、生活史、自我民族志(autoethnography)、女性主义民族志和民族志小说(Denzin, 1997; LeCompte, Preissle, & Tesch, 1993; Van Maanen, 1988)。这里重点介绍在社会科学研究中两种常用的形式:现实主义民族志和批判性民族志。

(1)现实主义民族志

这种形式的民族志是文化人类学家所使用的传统方法。Van Maanen (1998)认为,这种方法反应了研究者对研究对象所采用的一种独特的姿态。现实主义民族志客观的记录情景,通常以第三者的角度来客观地呈现在田野中获取的有关研究对象的信息。根据 Van Maanen (1998)的看法,在现实主义民族志中:

- 研究者以第三者的口吻讲述田野中的所听和所看,不必加入研究者的反省。研究者隐身幕后,如同一个全能者知晓并讲述一切"事实"。
- 以标准的方式汇报客观的资料,使报告不受个人偏见、政治目的和其他判断的影响。研究者提供所研究群落之日常生活的具体细节,并用标准的范畴进行文化描述,如家庭生活、工作、社会网络、社会地位等等。
- 虽然民族志研究者也通过引用参与者的话来建构被研究者的观点,但研究者最终决定文化如何被呈现和解释。

(2)批判性民族志

当 Denzin (1997)在谈代表性和合法性危机的时候,他其实是在对美国复杂的社会变迁做出回应。美国正变得越来越多元,正融入一个世界性经济,越来越多的族群成为美国人,诸如此类。这些因素产生了一个由权力、声望、特权和权威支配的社会体系,把不同阶级、族群和性别的人士边缘化。同时,20 世纪 20 年代德国批判理论所关注的一些历史问题,诸如压迫、异化和社会斗争,在当代的社会科学研究中已经难见踪影。

基于此,民族志现在已经发展出"批判"的取向(Carspecken, 1995; Carspecken & Apple, 1992; Thomas, 1993)。在批判性民族志中,作者主张解放社会中的边缘群体(Thomas, 1993)。典型的批判研究者具有明确的政治抱负,试图通过他们的研究来反对不平等和压迫(Carspecken & Apple, 1992)。例如,批判性民族志可能会去研究学校为某类学生提供特权,研究学校为不同社会阶级的子弟给与不平等的机会,或者研究教育体系如何使传统的课堂模式永久化,鼓励男生多发言而让女生沉默。

批判性民族志包括这样一些主要内容:价值导向,使人们能够得到更多的权利,挑战社会现状,关注权力与控制。以下是更为具体的特点(Denzin, 1997):

- 批判性民族志研究权力、不平等、不公正、统治、压迫、霸权和欺骗等社会问题。
- 研究者进行批判性民族志研究是为了防止他们所研究的对象进一步边缘化。这样,研究者与被研究者相互合作,讨论最后的书面报告,小心地进入和离开田野,并回报被研究者。
- 批判性民族志自觉地意识到自己的解释立场,承认解释本身反映了研究者自己的历史和文化,因而这些解释只是暂时性的,有些时候甚至很成问题,需要依靠读者和参与者自己的体会来具体理解。
- 批判研究者把自己置身于文本之中,并反思自己的角色。他们并非停留在幕后,而是表明自己的偏见和价值,也标明文本中哪些观点是自己的,哪些是参与者的。他们从来不标榜自己是"客观"的研究者。
- 批判研究者这种非中立的姿态也意味着他们主张变革社会,让人们少受压迫,

免遭边缘化之苦。

- 最后,批判性民族志的研究报告将会是"凌乱的、多层次的和多种方法并呈的",充满了矛盾、无法估量的因素和张力。

4.7.2 民族志的关键要素

如同 Hammersley & Atkinson (1985), Wolcott (1987, 1994),以及 Fetterman (1989)所认为的那样,在研究之始,民族志工作者往往观察人们在日常生活中的互动,辨认一些普遍性的模式,如生命周期、事件和文化主题。"文化"是一个难以归类的无形无象的概念,但它绝不是"点缀"性的东西(Wolcott, 1987,p.41),而是人们日常生活的模式,研究者把这些模式冠之以"文化"。从人们的语言,行动研究者可以揣摩出存在于这群人中的文化。研究者要观察他们做什么(行为),说什么(语言),他们应该怎么做,而实际又如何做,这之间存在何种程度的差异,以及他们制作和使用什么(工具)。因此,民族志工作者会收集一些手工品和实物资料;寻找故事、仪式和神话;或者发掘文化主题。这类主题有很多种,Winthrop(1991)在《文化人类学词汇词典》(*Dictionary of Concepts in Cultural Authropology*)中列出了很多,Fetterman (1989)把这些主题归为结构和功能两类来指导对社会组织的研究。结构是指群体的社会结构或构成,如宗族。功能则是指群体成员之间的社会关系模式,这些模式有助于调节成员的行为。

为了找出这些模式,民族志工作者投身于田野进行深入的研究,这被称之为田野工作。他们通过观察、访谈和实物来收集资料,然后运用这些资料来描绘该群落的"文化规则"。民族志工作者应该对田野工作的相关问题非常敏锐,例如要判断谁能帮助他/她顺利进入田野,这种能为进入田野提供途径的人被称为守门员。还要选择若干关键的信息员,这些人熟悉该群落的情况,能够为研究者提供大量洞见,还能建议研究者去访谈更多相关的人。研究者也要考虑如何回报被访者,被访者提供了信息,研究者也得有所表示才行。

研究者也应该敏锐地感知他/她自己的存在对被研究者生活的影响。根据研究伦理,研究者应该明示他的研究目的和意图,不要有任何欺骗。民族志研究提倡具体描述具有共享文化的群体,透过某些主题或视角来分析这些群体,解释这些群体中社会互动的意义并抽象出人类社会生活的一些普遍性规则(Wolcott,1994)。当然,在上述三个方面中,研究者可能会根据情况有所侧重。民族志研究的最终产品是一个有关被研究社会群体的完整的、文化的描述,其中包括了生活于其中的成员的观点,也有研究者从社会科学视角出发的解释。所谓完整的描述,就是民族志工作者尽量多地描述该群体生活的方方面面,比如历史、宗教、政治、经济和环境(Fetterman,1989)。所谓文化的描述,就是指民族志研究者需要概括地介绍该群体的文化全貌,把在田野中认识到的有关这一群体的各个面向都描述出来,展示其复杂性。

4.7.3 民族志的研究程序

如同所有的定性研究方法一样,尽管没有一个单一的方法来指导如何进行民族志

研究,但最近的一些研究已经提供了更为详细的指导(Wolcott, 1999,很好地介绍了这些方法):

1)研究一个文化共享的群体。研究的第一步,也是最重要的一步,就是搞清楚你为什么要做这一研究,准备采用何种研究设计,以及研究意图如何与所要研究的问题相关联。在所有形式的民族志研究中,这些问题都需要回答。研究的意图和你想回答之问题的类型在现实主义民族志和批判性民族志中大相径庭。对于现实主义民族志,研究的重点在于理解一个文化共享群体,然后从这个群体出发,得出一个有关某文化议题的深度理解。这一文化共享群体可以是一个学校,也可以是一个班级。至于文化议题则可能包括对某种文化的适应、文化传入、社会化、制度化教育、学习与认知、儿童与成年人的发展(LeCompte et al. ,1993)。在批判性民族志中,其研究意图与现实主义民族志大不相同。批判性民族志倾向于用他们的研究来主张变革,所以常选择一些具体的议题,如不平等、统治、压迫和解放等来研究。

2)制订一个计划来研究这一群体。因为民族志工作者往往会很长时间与被研究者在一起,取得他们的信任,找一个人做"守门员"以便进入田野,尊重当地的人与事,并对参与研究的人有所回报,这些都是民族志研究的重要内容。在研究之初,研究者需要确定哪些人、哪些地方有助于理解他所要研究的文化议题,或有助于人类解放。

3)询问有关行为、信仰和语言的共享模式的问题。民族志的研究问题往往探讨群体的某一文化面向,如"帮会成员使用何种风格的语言?",又如,"中学里的小混混(punkers)有哪些行为?"

4)描述这一文化共享群体并分析资料所展示的文化主题。在所有形式的民族志中,研究者都要进行概要的描述、分析资料,以及解释资料所揭示的意义。这些都是定性研究中常见的程序。不过,在不同形式的民族志中,具体的程序有所区别。在现实主义民族志中,研究者必须考虑描述、分析和解释之间的平衡,以便使每一部分都在整个作品中显得重要。而且,研究者需要解释现有文献对他/她所研究领域的现状,同时还要指出他/她自己的研究对这一领域的贡献。在批判性民族志中,描述、分析和解释都受"批判"内容的影响。研究者需要从主张变革的角度来解释其发现,这些变革被认为是有助于改善被研究者生活的,因而也是应该出现的。批判性民族志工作者经常提出一个具体的改革计划。

4.8 个案研究

尽管文化共享群体或具体的个人会被视为一个"个案",但是定性研究中的个案研究法不那么关注群体的模式,而更在乎对一个过程、事件或行动的深度描述。有人认为"个案"只是研究对象(Stake, 1995),也有人认为它是一种研究方法(Merriam, 1988)。不过这两种观点都认为个案研究是通过在多种信息来源和丰富的背景中收集深入、详细的资料,对一个或多个个案进行历时性的探讨。

社会科学家对个案研究并不陌生,因为它被广泛用于心理学、医学、法学和政治学。在很多学科里,个案研究都有悠久而显著的历史地位。Hamel (1993)把个案研究在社会科学里的源头追溯到人类学和社会学。他认为,人类学家 Mallinowski 对 Trobriand 岛的研究,法国社会学家 LePlay 对家庭的研究以及芝加哥大学社会学系进行的个案研究,如 Thomas & Znaniecki 的《在美国和欧洲的波兰农民》(*The Polish Peasant in Europe and America*),是定性个案研究的先驱。现在,研究者可以有很多选择去进行个案研究。例如,Yin (1989)同时用定量和定性的方法来进行个案研究,并讨论解释性和描述性的定性个案研究。Merriam (1998)在教育学中主张普遍使用定性个案研究法。社会学家 Hamel 从历史和问题导向的角度对定性研究方法进行讨论。Stake (1995)系统地提出了个案研究的程序,并在他所提供的"哈珀学院"(Harper School)例子中广泛运用。

4.8.1 不同的取向

个案研究法与众不同且本身具有相当之价值,所以很多学者选择这种方法。当个案本身很有意思,它就被称为内在性个案。Stake 对双语学校的研究是这类个案研究的典范(Stake,2000)。另外,定性个案的重点可以放在具体的议题上,用一个个案解释另一个个案。此种类型的个案会被视为一种工具性个案,因为它被用于说明某一特殊的议题。Asmussen 和 Creswell 描述一个校园的情况,并用这个工具性个案来描绘校园对枪击事件的反应。

个案研究也可以包括多个个案,这类个案研究被称为集群个案研究(collective case studies,Stake, 2000)。通过对多个个案的描述和比较可以用来深入探讨某一问题。比如,研究者可以比较几个学校,看看他们为学生提供了哪些不同选择。最后,研究者可以利用多个案例来分析一个具体问题。这样做的时候,研究者需要有意识地选择若干典型个案或代表性个案。虽然定性研究者无意将其发现普遍化,但也经常通过对若干个案的研究得出某些普遍性的认识。因此对典型个案的研究尤其重要。

4.8.2 个案研究的关键要素

个案研究中的个案是一个有限系统,受限于时间和地点。个案可以是一个计划、事件、行动或个人。例如,研究者可以选择多个计划(多地点研究)或单个计划(单地点研究)来进行研究。个案也可以是计划、事件或行动中的某一个人或多个人,如,实施新教学计划中的一个或几个老师。这些个案要能够反映事件或行动的过程。为了了解这些事件,研究者需要通过多种方式收集资料,如观察法、访谈法、音频和视频资料、文献和报告等。研究者试图通过收集多种形式的资料,如照片、剪贴簿、录音带和电子邮件,来对个案进行深度描述。

提供一个深度理解只需要少量个案,因为研究者不可能对很多个案都进行深度研究,没有这么多时间。研究时需要交待个案的背景,如当时社会、历史和经济情景。研究者也要把个案置于更大的背景中来理解,如地理、政治、社会和经济条件,例如,一个大家庭包括祖父、兄弟姊妹和其他被"收养"的家庭成员。

4.8.3 个案研究的程序

进行个案研究有多种程序可供选择(Merriam, 1998; Stake,1995)。在这里,我们主要讨论 Stake (1995)所提出的研究程序。

1)对一个有限系统进行深度研究。在运用个案研究时,研究者必须对深度讨论和有限系统的分析感兴趣。研究者需要判断:对研究而言,哪一类型的个案信息最丰富?个案本身是否重要?

2)提出有关的研究问题,并询问具有特殊意义之个案的具体细节。研究问题可能涉及一个案例或议题。例如,研究者会问:“对五个学校的老师来说,专业化意味着什么?”在这个例子中,焦点是以这五个学校为个案来了解某一具体议题。研究者同样可以问:“针对聋哑学生的远程教育的过程是什么?”这一个案研究的焦点是了解一个具有内在价值的特殊案例。

3)收集多种形式的资料来进行深度理解。因为个案研究最重要的一个特征就是对个案进行深度描述,所以研究者需要收集多种形式的资料,包括访谈、观察、文献和声频视频资料。资料的广度和复杂度把个案研究与其他定性研究方法区别开来。例如,Yin (1989)推荐了六种类型的信息:文献资料、档案记录、访谈、直接观察、参与观察以及实物资料。

4)详细描述个案并分析个案所反应的主题。对个案进行详细描述是数据分析的应有之义。研究者应该按时间顺序来陈述个案,也应该涵盖个案所涉及的人物、地点和活动。进一步的数据分析需要呈现个案所反映的主题,这些主题增加了个案分析的复杂性。在多个案分析中,研究者常把蕴涵于每个个案中的主题找出来,然后比较这些主题有何相同点或差异。这一程序被称为个案内和个案间主题展开。这种分析数据的方式可以整体性地分析整个个案,也可以部分分析个案的某一方面。

5)无论是描述个案还是论述观点,都要把他们置于其背景之中。这样做有助于把个案中具体的活动和情形与分析结合起来。分析的时候,重点要放在组织、日常活动、个案所涉及的人物和地点,以及相关的信息上。有了这些丰富的背景交代,个案就可以自我呈现(Merriam, 1988)。研究者也可以运用一些技巧来进行叙述,这些技巧包括年代学,对主要事件的叙述以及对个别事件的详细描述。

6)解释个案的意义。研究者需要对个案的意义做出解释,这种意义或者来自工具性个案所蕴涵的主题,或者源自内在性个案不寻常的情形。如 Lincoln 和 Guba (1985)所建议的那样,研究者在最后的解释阶段需要报告他/她从个案中“学到的经验”。

4.9 定性研究方法论:一个新的视角

在这部分的介绍中,我们讨论了当研究者对有关现象的真实信息感兴趣而又不满足于方便的定量方法所能获得的信息时,定性研究方法如何被用于收集资料。下面这

篇被允许引用的文章将讨论这些新的研究形式是如何出现的,它们的前景以及待解决的问题。这是一次有意思的回顾和展望。

定性研究方法论的新领域

Elliot W. Eisner

首先,我要感谢定性研究特别兴趣小组(简称SIG)的成员在SIG十周年庆典之际邀请我做这次演讲。十年前,当SIG在华盛顿举行第一次学术会议时,美国教育学研究学会(简称AERA)的成员应者寥寥。不过从那时起,定性研究方法,以及整个教育研究在这条路上一路行来,渐行渐强。在座各位都见证了这一过程,并为此做出了贡献。在此,我对各位的成就表示祝贺,也借此机会反思一下我们今天的处境。

我主要想讲讲定性研究方法新领域的前景和待解决的问题。何谓新领域?为什么它现在会被关注?它将对提高教育质量有何帮助?有哪些问题需要解决?最后,我们要向何处去?这些都是我将要讨论的问题。

定性研究方法之新领域涉及以下内容:关于认知的新假设、研究的意义,以及新方法会如何拓宽或补充传统教育研究的思考方式。传统的教育研究方法,也即统计法、相关分析和实验法,为如何操作研究提供了范式。在过去,做研究就意味着使用这些方法。被一整套技术语言和动力武装着,传统研究方法试图从数据中得出可靠的结论。这些研究方法及其假设很健康也很有生命力,但对教育研究而言,它们不再是排他性的,甚至不再是占优势的研究方法。为什么会出现这一情形?

我所称的“新领域”的出现得益于若干因素,其中之一就是试图摆脱操作主义、实证主义和行为主义的束缚。受那些传统方法支配的研究往往过于化约。在很多研究者看来,这些研究所忽略的内容比他们所揭示的还要多。例如,操作主义要求测量变量。虽然测量本身不失为一种描述世界某些方面的精确方法,但是它也牺牲了对很多细节的陈述。越来越多的研究者开始意识到形式与内容不能割裂,意识到一个人选择何种方式进行陈述会影响到其陈述的内容。

至于行为主义,无论是根植于美国心理学家桑戴克之研究的刺激反应理论,还是华生的理论都不能充分理解世界如何作用于人们,尤其是儿童,也不能理解世界对他们有何意义。人类生活在一个充满偶然性的世界,其动机也会随变化的情景而转移或变化。的确,在理解人们如何赋予这个社会以意义时,建构主义变得越来越重要,而理解个体心理的行为主义的声音则越来越弱。在很多教育学者看来,20世纪前五十年里占统治地位的哲学和心理学取向已经过时了。这种看法事关重大。

对传统方法不满意的另外一个原因是人们对Schwab所称的实践越来越感兴趣。在他那篇经典论文中,Schwab提出了一种对知识的新看法。那篇文章是在1969年的年会上宣读的,可惜当时只有少数几个教育研究者意识到它的价值。对于Schwab来说,教学和课程发展都是实践活动,而不是别的什么。他是在亚里士多德的意义上使用实践这一概念的:实践活动就是要做出好的抉择而不是寻找真理;它涉及偶然性而不是因果关系;实践活动运用理论,遵循规则,但不是规则本身;实践活动需要深思熟

虑,并尽可能地运用古希腊人所谓的实践智慧,实践智慧不仅涵盖知识也包括明智的道德选择。

Schwab 对实践的强调为教育研究者预示了另一种思考途径。在座熟悉亚里士多德的人都知道,亚里士多德区分了三种类型的知识。第一种是理论性的知识,这种知识原则上是关于那些具有必然性的现象,这些现象按惯常的方式出现,绝对不会以别的方式出现。在数学、天文、物理等学科中,其研究对象在条件充分和必备时,是可以被充分认知的,而且如亚里士多德所言,这些对象的运动轨迹内在于其自身。

我所说的实践性知识主要是偶然性的知识,它取决于具体情景。或许我们可以从政治艺术中窥见实践性知识的特征。借用一下亚里士多德的语言来说,我们显然不能用理解行星运动的方式来理解人类的行动。

第三种知识的形式是生产性的知识。亚里士多德认为,生产性的知识就是关于如何制作东西,诸如桌子、乐曲、绘画和诗歌。艺术最能体现生产性的知识。

从 Schwab 的文章中我们可以得出这样的认识:我们不能简单地用实证主义的概念来定义知识,然而实证主义如此长久地支配了教育研究,这样生产出来的理论在实践领域用处不大。理论是普遍性的,但行动是特殊性的。实践性的知识告诉我们具体该做什么,而理论性的知识仅提供背景情况。

大约在 Schwab 陈述这篇论文的前后,学界开始对认知多元论越来越有兴趣。这种兴趣受到哲学发展的刺激,如 Rorty 的著作,也受到女性主义和多元文化主义的影响。在女性主义者看来,传统所谓"可靠"且"严格"的数据都是以男性的标准来衡量,因此这些数据对女性主义者没有太大的吸引力。更多的内容有待我们去理解,而不是坚持男性视角,抱残守缺。对传统研究方法的不满意和对新方法的兴趣促使学界回归到对个案以及特殊性的研究。在教育研究领域,Jackson (1990)的《课堂生活》(*Life in Class Rooms*)和 Smith (1968)的《城市教室的复杂性》(*The Complexities of the Urban Classrooms*),开了这方面研究的先河。

至少在最初,一些传统的研究者认为定性研究仅仅是为"真正"的研究铺路,尽管他们对个案研究也会感兴趣。不过他们会问,除了解释个案自身,我们能从个案中获得什么知识呢?回答这个问题需要了解一下历史。事实上,对个案进行研究的历史源远流长。Toulmin 清楚地指出在 17 世纪中期现代科学到来之前,人们早就对特殊性感兴趣。伽利略对关联的定量描述改变了我们对科学是什么的认识。这种观念上的转变让人们关注永恒的而非及时的,普适性的而非地方性的,文字的而非口述的,一般的而非特殊的,定量的而非定性的。Toulmin 认为到 20 世纪 40 年代末和 50 年代,人们又重新对及时性的、地方性的、特殊的和口述的知识感兴趣。的确如此。

我以为,这些新的兴趣在以下情形中得到了公开的表达:定性研究特别兴趣组的会员越来越多,定性研究的学术刊物和著作不断涌现。去看看美国教育学会年会的论文目录,你就会发现定性研究的论文非常之多。或许关于诸如教师培训和学校改革之类的论文会比定性研究的论文要多,但是即使这些领域学者也开始关注及时性的层面。以下的数字也告诉我们定性研究变得越来越重要。在 1991 年的美国教育学年会上有 22 篇文章涉及定性研究,1992 年为 32 篇,1993 年为 18 篇,1994 年为 27 篇,1995

年为50篇,1996年为55篇,今年是45篇。

我曾经说过,新领域的出现是源自对旧范式的不满,对认知多元论的兴趣以及对研究之本质的反思。新研究路径之一就是晚近出现的以艺术为基础的研究。如各位所知,在定性研究被教育学者视为一种可行的和可定义的方法时,最初关注的是民族志。这是可以理解的。毕竟,民族志是人类学的产物,而人类学是社会科学的一种。从一种社会科学转向另一种社会科学,相对而言,比转向人文取向的研究路径要少一些阵痛。那种认为艺术能为研究提供基石的想法曾被很多人视为自相矛盾。然而学者们逐渐认识到,科学的研究只是研究的一种类型而已。研究并非是社会科学的专利。事实上,任何细致的、反思的、系统的和有助于深化人们理解现象的探讨都可以被视为研究。关键在于如何去做。

我想要说的是,美国教育研究学会正处于一个特殊的历史时期:解释取向愈发显著,我们正见证历史。尤其是年轻学者,他们正在创造新的研究方法来探讨他们所关心的教育问题,但是他们并不排斥其他方法。以艺术为基础的研究就是新方法的一种。我很高兴地告诉大家,在由美国教育学会资助的三个从事以艺术为基础的研究的研究机构中,过去四年都招募到超出预期人数的会员。

到目前为止,我的评论只是想回顾一下美国教育研究领域内的定性研究。但我也得告诉大家,对定性研究感兴趣的绝非只有教育研究领域,社会科学家和其他领域的学者同样对此感兴趣。例如,人类学家 Schepper-Hughes (1992),社会学家 Becker (1990),心理学家 Sternberg (1988)。如我曾说,这是一个动荡的年代。

现在我们把关注点转移一下,看看定性研究领域里之新方法的前景。这些新方法必须提供些什么呢?

教育研究之新领域发展的可能结果之一就是它流行至无形。新领域的出现使传统的研究观暴露出问题,让人们重新审视过去想当然的假设。如果世界都是灰色的,那么我们将看不见任何东西。正是由于对比——定性的对比或更常见的概念对比——才使我们得以分辨熟悉的事物。当每个人都在量化这个世界时,除了量化,我们似乎没有别的选择。研究者被要求从总体中抽取个案然后推论总体,这些做法被视为天条,不容置疑!当每个研究者都根据教条化的条件来推论效度时,效度的意义就在于它变成了新手必须记忆的教条。

新的研究方法以及循此生产的知识开启了争论。我自己曾经经历过的一次学术争论发生在1996年的教育学年会上,当时我与 Howard Gardner 争论小说是否可以作为博士论文的一种形式。Gardner 持反对的观点,而我持支持的观点。当然,我们持不同的观点并非出于论战的需要而是真诚地这样认为。对这一议题进行争论,这本身就意义重大。十年以前谁会想到小说会被考虑成为博士论文的一种合法形式呢?(不过我得告诉大家,我的一位杰出的同事,也是世界顶尖的组织理论家,James G. March 对我说,在他看来,大多数美国教育学所做的研究都是小说)替代性研究方法的存在迫使我们为其寻找理由,这样做也使我们明白了传统方法的用途和局限。从某种程度上说,我们的自觉性提高了,因此我们更可能了解正在做的事情。无疑,这是一个进步。

新领域的第二个可能的发展方向则是对叙事的运用和探讨。或许大家都知道,Bruner (1985) 区分了认识的例证模式和认识的叙事模式。前者寻找真理,后者寻找逼真或者类似真理的观察。在前者那里,你所指即是你所说;在后者那里,你所指超过你所说。叙事涉及讲故事和经验分享。体验本身被视为教育的首要媒介,于是故事就成为分享体验的最常用的方式。叙事,或者说讲述,使他人分享你的或其他人的体验成为可能。如果运用得好,叙事可以激发想象。通过想象性的参与,我们能更加清晰地建构出"真实世界"。更为重要的是,当叙事被很好地运用时,移情式的理解得到发展。

请允许我对移情多讲几句。传统的观点不把感情当作理解的对象。感情充沛似乎意味着失去控制。理性是冷酷无情的,感情无缘置喙。的确,启蒙运动一个很棘手的遗产就是身体与心灵的分离。这种分离在学校教育以及我们关于人类能力的理解中仍然存在。叙事研究所做的,就是把我们置于情景中,帮助我们理解人们的感受。比如一个 46 岁的副教授被妻子唠唠叨叨地抱怨他没有出息时,这位老兄在大学里以及生活中的感受。这就是 Albee (1962) 在《谁怕维基里亚·沃尔夫》(*Who's Afraid of Virginia Woolf?*) 一书中展示给我们的景象。Mailer 的《晚上的军队》(*The Armies of the Night*, 1968) 和 Capote 的《冷血之中》(*In Cold Blood*, 1965) 让我们得以窥见战争和谋杀的世界。叙事能够"到达"(get at) 数字和文本不能企及的境界。请注意,我用的是"到达",这是我刻意选择的字眼。

我相信,在理解学校和课堂、家庭和社会,以及教学实践和管理时,我们既可以使用叙事研究的方法来进行,也可以用诸如相关系数、T 检验之类的定量研究词汇来理解。在此,我声明我并不反对用统计法来进行研究,但我的确不认同那种把统计视为惟一合法方法的观点。

在此,我想要说的是,在人类有能力进行沟通之时,故事就已成为最主要的沟通方式。在某种意义上,民族志只不过优化了这些故事,而好的艺术则是把讲故事的形式和内容很好结合起来的精华之作。

新领域的第三个可能后果是给那些受雇于学术机构的人更多机会。我想说,方法与智力相互作用,并非每一种方法都适用于每一个研究者。定性研究方法的出现为研究者提供了更多的机会,让他们能够选择一种研究方法。这种方法最适合他们的研究兴趣,与他们的研究目的相匹配,能展示他们的实力,也能很好地利用他们的智力,并给他们一个机会和公开的舞台来做到这些。

当然,也有一些研究者满足于并乐于做统计研究。对他们来说,这些方法更好。不过我在这里所鼓吹的不是把新领域塑造成新的学术霸权,而是拓展可用于研究的资源。这些资源过去不被看好,而现在,通过可行的研究程序并利用语言和象征的方式来表达,它们正变成可供研究生和教授利用的材料。如此一来,研究方法得到了丰富。这一过程正在进行,相信在座各位会乐观其成,我也如此。

新领域的第四个可能后果是合作的观念。它是随之而来的,但并非必然如此。女性主义者最早提醒我们关注在教育研究中研究者与教师(被研究者)的彼此对立,这种对立会影响到我们对真实情形的理解。很多团体督促我们不要把教师和学校管理层视为研究对象。他们是研究的合作者,同研究者在一条船上。虽然各自承担不同的

职责,但都能做出不可替代的贡献。当这种合作发生时,研究者就更加可能用人类学的概念和移情理解来生产内幕性的知识。而且,这类合作不仅发生在资料收集阶段,也会在概念化过程中出现。

真正的合作,在我看来,研究者和教师仅仅有良好的愿望还不够。它需要重新界定教师的角色以便教师能有更多的机会和时间参与到合作研究中来。诚然,学校是为学生解惑的地方,不过学校也不妨成为解答老师疑问的场所。

以上我描述的这些前景只是存在于可能之境。下面我将回答:新领域可能遇到的问题是什么?我们需要注意什么?

在谈到实质性问题之前,有一点我想说明:下面我要讲的内容,有一些比较成熟,但也有一些你们可能会认为很糟糕。不管怎么说,这些都是我的真实想法,我愿意在今晚同你们分享。

坦率地说,对于新领域,我不持过分乐观的态度,但也不气馁。我将列出我尚不能解决的问题,以及我个人认为对定性研究发展前景至关重要的问题。我希望你们能认真考虑并解决这些问题。总之,我的呈述是建设性的。

让我列出一些我自己面临的实实在在的难题。最明显的就是很难出版非文本性的材料。几年以前,在就任美国教育学年会主席一职的演讲中,我曾经播出过一段录像来演示非语言形式的认知功能。当这篇演讲稿后来刊登在《教育研究者》(*Educational Researcher*)上时,可以想象,文中的个案无法以文本形式呈现,精华被剥离了。当我从电影和音乐中截取片断用于研究时,也出现了同样的困境。学术刊物没有容纳这些材料的机制。或许未来可以用 CD-ROMs 来解决这个问题,但眼下还无法解决。

另一个实际的问题就是大多数大学很少或者根本没有运用这些新方法的传统。因此,很少有老师能指导学生做这方面的研究。这种情形很普遍。从定义上讲,创新意味着新,不过新也意味着经验上受到局限。那些把电影作为主要研究手段的人不仅需要有机会接触到摄录机,也需要了解如何制作电影。除非学校或者系里支持这一研究取向并提供相应的器材,要不然这类研究工作前景不妙:没有老师指导使学生缺乏使用相关媒介的技术。尽管其动机崇高,结果可能就是一场灾难。到最后,所谓创新就可能是把自己能企及的材料弄到一起,然后尽可能地把他们结合起来。这当然是一个亟待解决的问题,尤其是对大学而言。我在这里提到这一点是因为它的确影响到了研究质量。

与创新和经验有关的另一个话题就是:如何评价与传统范式不符的研究?当一份教育研究报告以小说的形式呈现时,该如何评价?应该由谁来评价?以什么标准?如何评价一个多媒体论文?信度与效度、概括性和稳定性,这些标准还适用吗?如果不适用,那应该是哪些标准呢?

或许我们能从艺术中学到某些东西。我们都知道梵高在世时仅卖出一幅画,而斯特拉文斯基的《春天的仪式》初次上演时台下观众嘘声一片。现在,新形式需要新标准,而新标准的出现则需要那些能够解释新作品意义的人来做大量的工作。艺术中的批评在艺术创作中就扮演了这样的角色:评价创新性作品。

除了这些实际存在的问题,还有一些理论性的问题。比如,定性研究者往往不愿

为他们的描述提供分析性或理论性的解释。他们的观点是艺术作品只代表它自身，一出戏剧的作者并不需要提供有关戏剧之于观众的理论意义。戏剧已经在语言、行动和舞台上充分地得到了展现，它已经包含了不可言说的意义，但观众是否领会就不得而知了。

以上这些用于戏剧的标准被类推到教育研究。对此，我存有疑问。我们的工作必须比一个好的记者或好的作家所能做到的工作更好。毕竟，人们期望我们能为教育研究提供一些理论性的认识和背景。现象需要在理论的关照下被推论、分析以及评估。正是这种理论背景为我们的工作增加了价值。

简单地说，我相信，除了极个别的情形，我们的同行很难满意那些不加任何解释的堆砌材料的定性研究作品。真正的挑战在于如何做到既有理论解释，又不牺牲材料本身的丰富性。据我所知，Jackson(1993)为我们提供了一个极好的范本。在他的论文中，Jackson 创造性地将活泼的叙述和哲学性的分析很好地结合了起来。我相信我们需要更多的此类作品，不过这还只是期望而已。我想再次强调，仅仅有好的，甚至是完美的描述都是不够的，理论性的分析不可或缺。

我另外一个关注点与定性研究的模糊性有关。在文学中，朦胧具有正面和建设性的意义。不过，朦胧也会导致不确定性，使读者无法清晰地了解故事所涉及的现象，进而没办法知道研究的意义。定性研究一方面要追求清晰，另一方面也要追求具有创造性的朦胧。我感觉这之间有一种张力。现在我还没有办法解决这个问题，只是想把这种两难困境指出来，让在座各位思考：如何具有想象力地研究教育问题，又不至于陷入罗尔沙赫氏测验(视对墨渍图案反应而分析其性格的实验)式的不确定性之中。

“站不住脚的相对主义”也是我所关心的。我用“站不住脚”一词来指这样的研究态度：既然解释都是个人性的，那么任何解释都是站得住脚的。这一态度的潜在逻辑就是：没有一个共同的标准来判断研究的质量和意义。尽管定性研究不再追求揭示“真理”，但是不能以此认为从此就没有一个共同的基础来评估研究。例如 Rorty 放弃了揭示真理，转而追求“启迪”。但他终归有所追求。换句话说，好的研究应该能丰富或更新我们的知识，或者至少要让我们看到它正在朝某个方向前进。但我很担忧作品本身的创新性和评价标准上的相对主义会使我们失去方向。该如何评价新领域里的研究？哪些标准合适？我们如何才能从沙砾中分辨出金子？如果理论不再与真理相联系，哪些评价标准仍然适合？以上这些都是我的一些关于定性研究评价标准的疑问。

造成这些困扰的关键在于，在定性研究中——其实定量研究也是如此——原始资料不为读者所知。出于实际原因，这类资料没有被存档。即使有些时候有存档，这些资料也与研究者的体验相割裂，变成一堆苍白且目的性很强的复述。基于这样的事实，定性研究的质量很大程度上取决于作者或电影制作者本人。

我当然知道三角测量(triangalation)、结构性验证、引用充分等方法可以用来检验待验证的结论或归纳。但是如何进行描述的底线掌握在研究者手中，我们无从得知这些描述的精确性。我相信我们需要反思这些问题。

与个人化解释相联系的另一个问题是研究的说服力。所有的研究，不管是定量还

是定性，都试图说服人们相信某些说法。广告和宣传同样试图让人们相信某些东西。我们该如何看待其说服力呢？哪些要素构成了定性研究说服力的基础呢？

关于这个问题，Barone和我在《教育研究的补充方法手册》(*Handbook of Complementary Methods in Educational Research*)中给出了一个答案。在关于以艺术为基础的研究这一章中，我们认为这类定性研究不企图让读者相信某些结论，而是提出一些有质量的研究问题。坦白地说，这个答案并不让人满意。能有什么办法把定性研究的说服性置于坚实证据的基础上，同时又避免化约主义？换句话说，我们是否能既资料充分同时又充满想象力和诗意地进行定性研究？这个问题很难回答，我也不认为自己已经给出了答案。这些正是困扰我的问题，我们这些做定性研究的人绝对不能忽略这些问题。如果我们忽略，那么，那些不同意我们研究方法的人会提醒我们注意。

让我来谈谈与相对主义相关的另一个问题："专家"的概念。大家都很清楚，我们生活在一个质疑专家的年代。某些专家，比如鉴赏家，在很多人看来他们还多多少少胜人一筹。但是，随着民主的趋势在研究机构的各个层次的发展，如果还有人认为某些人在解释和分析方面高人一筹的话，这种想法会被当作权威主义。我得承认我并不准备放弃专家或者权威的观念。根据社会学的概念，权威可以划分为先赋性权威(ascribed authority)和自致性权威(achieved authority)。前者是占据了科层制中的某个要害位置，后者通过能力和努力而达致。相对主义的评价把专家边缘化了，把对作品的评判被化约为"纯粹的"偏好或品味。我不认同这种做法。

我相信同其他领域一样，定性研究中的专家在特定领域的确比其他人更精通一些，比如Whyte(1943)对波士顿意大利移民有很深刻的理解，Peshkin (1986)很了解美国中西部原教旨主义，而Kozol (1968)对波士顿的课堂教学了如指掌。定性研究常被用来解释音乐或艺术教学的质量，这就要求研究者不仅懂音乐和艺术，还要了解教学。我发现有些研究在描述课堂教学时往往忽略了课堂内容，似乎教学内容和教学实践无关紧要。我想说，专家的确重要，一个好的从事博士研究的人应该成为他所在领域的专家，能够为相关的实践提供有用的指导。

现在我们回到有关推论的问题。现在定性研究中的一个问题就是把结论局限于个案本身。在做个案研究时，我们是简单地呈现异群体或亚文化，还是把从个案中所得出的结论推及其他个案？

在座各位都了解，统计研究有一套规范的程序来进行推论，这套程序要求先在总体中进行随机抽样，然后对样本进行研究，最后推论总体。整个程序逻辑清楚、滴水不漏。由于定性研究的对象是方便抽样的样本，而非随机样本，那么我们该如何看待定性研究中的推论？

在我看来，在教育学中，定性研究生产的是一系列观察和图像，当我们考察其他学校和课堂时，这些知识有助于我们理解相应的现象和过程。Powell, Cohen和Farrar (1985)在《商场化高中》(*The Shopping Mall High School*)中告诉我们，高中学生和老师之间的协议并不局限于特定的某个高中。他们为我们提供了一个分析框架，这个框架不仅有助于理解学校和课堂，还有助于了解为什么会达成此类协议。考虑到高中老师每天必须处理大量的事务，他们必须找到一个折衷的办法，在这样的背景下，我们就能

理解学生和老师为什么会达成这样的协议了。

Powell 和他的合作者提出的"协议"概念可以用来概括其他学校存在的类似的计划，也能让我们深刻地了解到这一现象：学生和老师之间的协商可以使生活变得可以忍受一点。

我们是否可以认为定性研究得出的启发性结论与定量研究的结论截然不同？非也。尽管统计推论的逻辑看似无懈可击，样本可以推论到总体，但却不能推论到其他总体。首先，研究者需要了解总体的特征，以便判断从中得出的结论是否能用于解释其他总体。但很少有研究者指出其解释的边界。我们把定量研究的结论类推到其他总体，而且不考虑时间因素。事实上，我们并不知道在 1994 年有关某一总体的结论能否用来解释 1997 年的类似总体。时间关系重大，当时被研究的总体可能与现在我们想要解释的总体在特征上不一样。简言之，定量研究在推论方面与定性研究并没有本质性的不同。人类不仅会使用工具，也会使用象征，还可以进行推论。我们总是试图从很多个案中选择一个进行研究，然后把结论做些调整以适合所有个案。这样做的时候，我们使用类比和隐喻的方法，来使这个过程显得有道理。

在这次讲演的最后，我将谈谈方法的新领域会对教育研究的未来有哪些影响和挑战。

在我看来，使用定性研究方法的趋势不可逆转，不可能再回到传统的定量方法占支配地位的年代，像 20 世纪 50 年代和 60 年代那样。教育研究中出现的定性研究趋势是方法多元化的一部分。不过我也不认为美国公众广泛地接受了我们的认识论取向。公众已经习惯于把学术实践看作赛马，而对于赛马来说，数字似乎比叙述更有效。换句话说，推力一旦作用，就会在惯性的作用下继续下去，学术评价也如此。

我相信，美国人习惯于用比较的方式来衡量学校教育的成果。但是当某些人具有比较特殊的资质时，比较就很难进行了。出于目的考虑，我们希望所有的学生在同一条跑道上赛跑，用同一种方法测量，每个人的表现可以精确到小数点后的三位数或四位数。类似方法仍然支配着学校教育，定性的评价方式很难发挥作用。另外，同定性研究相比，定量研究的劳动投入少而且精确性高。虽然数字提供的是虚假的安全感，但不管怎么说也是安全感。

不过我也看到，定性研究在学术界比以前更受欢迎。我经常在一个星期里看到四五本有关定性研究的著作出版的消息。学界广泛关注定性研究，即使那些传统取向的同事也开始留心定性研究。

我尊重那些传统取向的同事，我感觉，选择方法或研究取向不仅是方法论上的抉择，更反映了人格特质。有人曾经说，这个世界上的人可以划分两类：一类人喜欢旧金山，另一类人喜欢洛杉矶。我想说，世界也可以划分成这样两类，喜欢定性研究的和喜欢定量研究的。直白地讲，我希望看到一个差异而多元的方法世界，人们可以在同一研究里使用不同的方法。正如我早先讲的那样，我无意建立一个新的支配性方法，而是欢迎最大限度地将方法多元化。

今晚，在演讲之初，我对你们在这个领域的成就和领导地位表示了祝贺。我真诚地这样认为。今晚我试图指出新领域的发展前景和未来需要解决的问题。这些问题

既重要又有趣，我相信有一天终会得到解答。你们的工作正在为后来者铺平道路，让他们在从事定性研究时不像你们曾经经历过的那样艰难。开辟新领域不是一件容易的事，它充满风险，也需要勇气。持久性是生命力的应有之义，今晚我们共聚一堂庆祝我们的十周年，有什么能比这更能证明我们所从事工作的持久性呢？

为定性研究者推荐更多的资料 4.10

4.10.1 推荐读物

Anzul, Margaret, Evans, Judith F., King, Rita, & Tellier-Robinson, Dora. (2001). Moving beyond a deficit perspective with qualitative research methods. *Exceptional Children*, 67, 235-249.

Berg, Bruce Lawrence. (2000). *Qualitative research methods for the social sciences.* Needham Heights, MA: Allyn and Bacon.

Bernard, H. Russell. (2000). *Social research methods: Qualitative and quantitative approaches.* Thousand Oaks, CA: Sage.

Campbell, Kim Sydow. (1999). Collecting information: Qualitative research methods for solving workplace problems. *Technical Communication: Journal of the Society for Technical Communication*, 46, 532-545.

Creswell, John W (1994). *Research design: Qualitative and quantitative approaches.* Thousand Oaks, CA: Sage.

Dana, Nancy-Fichtman, & Dana, Thomas M. (1994, April). *Holistic perspectives on the teaching of qualitative research methods.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Denzin, Norman K., & Lincoln, Yvonna S. (1995). Transforming qualitative research methods. *Journal of Contemporary Ethnography*, 24, 349-358.

Frey, Lawrence R., Anderson, Shawny, & Friedman, Paul G. (1998). The status of instruction in qualitative communication research methods. *Communication Education*, 47, 246-260.

Krane, Vikki, Andersen, Mark B., & Strean, William B. (1997). Issues of qualitative research methods and presentation. *Journal of Sport and Exercise Psychology*, 19, 213-218.

Lavender, Tony, & Lake, Nick. (2000). Qualitative research methods in psychiatric rehabilitation. *Psychiatric Rehabilitation Skills*, 4, 321-339.

Lindlof, Thomas R. (1994). *Qualitative communications research methods.* Thousand Oaks, CA: Sage.

MacKay, Kathleen A., & Schuh, John H. (1991). Practical issues associated with qualitative research methods. *Journal of College Student Development*, 32,424-432.

McLean, Les, Myers, Margaret, Smillie, Carol, & Vaillancourt, Dale. (1997). Qualitative research methods: An essay review [Electronic version]. *Education Policy Analysis Archives*, 5(13). Retrieved August 28, 2001, from http://epaa. asu. edu/epaa/v5nl3/

Miller, Steven I., & Fredricks, Marcel. (1996). *Qualitative research methods: Social epistemology and practical inquiry.* New York: Peter Lang.

Morse, Janice M. (Ed.). (1993). *Critical issues in qualitative research methods.* Newbury Park, CA: Sage.

Page, Reba N. (1997). A thought about curriculum in qualitative research methods. *International Journal of Qualitative Studies in Education*, 10, 171-173.

Patton, Michael Quinn. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.

Robinson, Rhonda S., & Driscoll, Marcy P. (1993, January). *Qualitative research methods workshops: An introduction, definitions. Readings on qualitative research: An annotated bibliography.* Paper presented at the Association for Educational Communications and Technology Sponsored by the Research and Theory Division, New Orleans, LA.

Smithmier, Angela. (1996, June). *The "double bind" of re-presentation in qualitative research methods.* Paper presented at the Qualitative Research in Education Conference, St. Paul, MN.

Smith-Sebasto, N. J. (2000). Potential guidelines for conducting and reporting environmental education research: Qualitative methods of inquiry. *Environmental Education Research*, 6, 9-26.

Steen, Mangen. (1999). Qualitative research methods in cross-national settings. *International Journal of Social Research Methodology*, 2,109-124.

Strauss, Anselm L., & Corbin, Juliet M. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed.). Thousand Oaks, CA: Sage.

Taylor, Steven J., & Bogdan, Robert. (1998). *Introduction to qualitative research methods: A guidebook and resource.* New York: John Wiley and Sons.

Westbrook, Lynn. (1994). Qualitative research methods: A review of major stages, data analysis techniques, and quality controls. *Library and Information Science Research*, 16, 241-254.

Whitt, Elizabeth J. (1991). Artful science: A primer on qualitative research methods. *Journal of College Student Development*, 32,406-415.

4.10.2 学术机构或团体

Association for Qualitative Research at www. aqrp. co. uk/

ES R C Qualitative Data Archival Resource Centre at www. essex. ac. uk/qualidata/
International Institute for Qualitative Methodology at www. ualberta. ca/ ~ iiqm/
Qualitative Research Consultants Association at www. qrca. org/
Qualitative Research Interest Group (QRIG) at www. upa. pdx. edu/QRIG/
Society for the Study of Symbolic Interaction (http://sun. soci. niu. edu/-sssi/) is a social science
professional organization of scholars interested in qualitative, especially interactionist, research.

4.10.3 定性研究的 E-mail 讨论小组

讨论小组或者电子公告及邮件交流(discussion groups or bulletin boards and listservs)作为学术交流的工具其作用日益显现。来自 Georgia 大学的 Judith Pueissle 教授整理出了一个关于此类讨论小组的名单,并提供了如何加入的指导。此处的使用获得了授权。

QUALRS-L@ listserv. uga. edu: Qualitative Research for the Human Sciences. To subscribe, send this message to listserv@ listserv. uga. edu: subscribe QUALRS-L yourname. For help, contact Judith Preissle, jude@ arches. uga. edu, or check http://listserv, uga. edu/archives/qualrs-l. html

QUALNET@ listserv. bc. edu: Qualitative Research in Management and Organization Studies. To subscribe, send this message to majordomo@ listserv. bc. edu: subscribe qualnet. For help, contact Ted Gaiser, Gaiser@ bcvms. bc. edu

ATLAS-TI@ atlasti. de: Topics on the text analysis, text management, and theory building program ATLAS/ti. To subscribe, send a on-line message to listserv@ atlasti. DE: SUB ATLAS-TI yourfirstname yourlasmame your institution. For help, contact Thomas Muhr, muhr@ atlasti. de

QUAL-SOFTWARE@ jiscmail. ac. uk: a list on qualitative analysis computer programs. To subscribe, send this message to jiscmail@ jiscmail. ac. uk: join qual-software yourname. For help, contact qual-software-request@ jiscmail. ac. uk

QSR-Forum@ qsr. com. au (Qualitative Solutions and Research), for the qualitative analysis programs NUD * IST and NVIVO. To subscribe, send a message to mailing-list-request@ qsr. com. au with the words SUBSCRIBE QSR-FORUM in the main body of the text. For help, contact listmaster@ qsr, com. au

QUAL-L@ scu. edu. au: Qualitative Research List, initiated to serve folks at Penn State, but immediately attracted a broader audience. To subscribe, send this message to listproc @ scu. edu. au: subscribe QUAL-L firstname lastname. For help, contact Bob Dick, bdick@ scu. edu. au

---

资料来源: Judith Preissle, Georgia 大学, 未出版,使用获得授权。

4.10.4 社会科学定性研究方面的网络资源

METHODS@ cios. org: a list for social science research methods instructors. To subscribe, send this message to comserve@ cios. org: join methods yourname. For help, contact support@ cios. org

PSYCH-NARRATIVE@ massey. ac. nz: a discussion of narrative in everyday life. To subscribe, send this message to majordomo@ massey, ac. nz: subscribe psych-narrative. For help, contact Andy Lock, A. J. Lock@ massey. ac. nz

Ethno@ cios. org: Ethnomethodology/conversation analysis. To subscribe, send this message to comserve@ cios. org: join ethno yourname. For help, contact support@ cios. org

IVSA@ pdomain. uwindsor. ca: International Visual Sociology Association. To subscribe, send this message to listserv@ pdomain. uwindsor. ca: subscribe ivsa yourname. For help, contact Dr. Veronika Mogyorody, mogy@ uwindsor, ca

VISCOM@ listserv. temple. edu: Visual Communications Discussion List. To subscribe, send this message to listserv@ listserv. temple. edu: subscribe viscom yourname. For help, contact Jay Ruby, v5293e@ vm. temple. edu

EVALTALK@ bama. ua. edu: American Evaluation Association Discussion List. To subscribe, send this message to listserv@ bama. ua. edu: subscribe evaltalk yourname. For help, contact ownerevaltalk@ bama. ua. edu

ARLIST-L@ scu. edu. au: Action Research Mailing List. To subscribe, send this message to listproc@ scu. edu. au: subscribe ARLIST-L Lastname. For help, contact Bob Dick, bdick@ scu. edu. au

XTAR@ listproc. appstate. edu: Teacher Researchers list. To subscribe, send this message to listproc@ listserv. appstate. edu: subscribe xtar yourname. For help, contact blantonwe@ conrad. appstate. edu, jmm@ am. appstate. edu, or combstm@ conrad. appstate. edu

GOVTEVAL@ nasionet. net: Public Sector Program Evaluation. To subscribe, send this message to majordomo@ nasionet. net: subscribe govteval [your email address]. For help, contact Dr. Arunaselam Rasappan, artd@ ppp. nasionet. net

H-ORALHIST@ h-net. msu. edu: H-Net/Oral History Association Discussion List on Oral History. To subscribe, send this message to listserv@ h-net. msu. edu: subscribe H-ORALHIST firstname lastname affiliation. For help, contact Jeff Charnley, Charnle2@ pilot. msu. edu

BIOG-METHODS@ mailbase. ac. uk: Biographical Methods for the Social Sciences. To subscribe, send this message to mailbase@ mailbase. ac. uk: join BIOG-METHODS lastname. For help, contact biog-methods-request@ mailbase. ac. uk

AELACTION@ ael. org: Appalachia Educational discussion list on action research in classrooms. To subscribe, send this message to majordomo@ aelliot. ael. org: subscribe aelaction your E-mail address. For help, contact owner-aelaction@ aelliot. ael. org

Q-METHOD@ listserv. kent. edu: Q Methodology discussion list on this broad approach to the study of subjectivity. To subscribe, send this message to listserv@ listserv. kent. edu: subscribe Q-METHOD yourname. For help, contact Q-Method-request@ listserv. kent. edu

ARMNET-L@ scu. edu. au: Action research methodology network. To subscribe, send this message to listproc@ scu. edu. au: subscribe ARMNET-L firstname lasmame. For help, contact Bob Dick, bdick@ scu. edu. au

Ethnography-in-education@ mailbase. ac. uk: use of ethnographic research methods in education. To subscribe, send this message to mailbase@ mailbase. ac. uk: join ethnography-in-education firstname lastname. For help, contact ethnography-in-education-request @ mailbase. ac. uk

OnlineRsch@ onelist. com: discussion of ethics and methodology in online research, including sociology, anthropology, and other related disciplines. To subscribe, send a message to OnlineRschsubscribe@ onelist. com; nothing necessary in the body of the message. For help, contact OnlineRsch-owner@ onelist. com

5

GUIDES TO METHODS AND TECHNIQUES OF COLLECTING DATA IN LIBRARY, FIELD, AND LABORATORY SETTINGS: SOCIAL SCIENCE DATA LIBRARIES AND RESEARCH CENTERS

图书馆、实地和实验室背景中收集数据的方法与技术：社会科学数据库及研究中心

要做一项良好的研究设计，数据收集始终是至为关键的一个步骤，因为一项研究的好坏往往有赖于所得数据的质量。在本书的这一部分，我们按照社会研究的常用领域，即图书馆、实地和实验室来逐一说明社会研究的方法和技术。我们谈到了一些主要方法的长处与短处，也对诸如如何设计问卷、如何安排面对面访谈、如何确定测量级别等问题提供了参考意见。

社会科学研究者的场地、主要方法与具体技术 5.1

在一份设计好的研究中，一般要在下列步骤完成之后再进行数据收集：

- 确定一个研究问题。
- 建构理论框架。
- 陈述理论假设。
- 设计具体问题。
- 决定抽样步骤。

这一节将介绍社会科学研究中最常用的一些方法，并简要说明如何设计问卷、如何当面谈访、如何编制量表，这些可以帮助研究者针对自己的特殊研究问题，来评估与选择较为合适的研究方法。同时，读者也应该较为全面地参阅本节所列出的专题书目，以求完整把握每一种方法或技术。

在我们开始列出书目表、讨论具体技术和指出资料来源之前，先谈谈研究方法，它与问题设计密切相关。一般说来，严格地区分这四种概念，即方法论、研究场所、研究方法、研究技术，有着重要意义。

方法论（相对数据收集而言）是一套描述、分析具体研究方法的知识体系，它说明研究方法的起源、局限性，阐明其前提条件及步骤、手段，揭示其对于研究所具备的潜在意义。在本书里，我们把重点放在详细阐明每种方法的优缺点上面，目的就是让读者明确收集数据的各种过程，然后再去选择较为有利的方法。

研究场所是指收集数据的具体地点，对于大多数学科来说，用得最多的是图书馆、实地以及实验室。

研究方法则是指所有学科或者部分学科会用到的，较为普遍的特定数据收集方式，它包括这样一些过程，比如：观察、测量、做实验、建立某种模型或理论、提供解释、作出预测等等。社会科学中常常会用到文献分析、问卷、个人访谈等方法。

研究技术是指一个具体研究方法中会用到的特定手段。例如，实地研究者会用社会测量表格去测量一些社会变量；人口统计学家会极大地依赖数据资料，运用各种数据统计方法去描述某种关系，或者对数据进行综合把握。

谈到研究方法与研究者具体的研究内容的关系，我们可以参考 Muzafer Sherif（1967）提供的、较为明确的建议：

> 就其本身而言，没有哪种收集数据的手段或技术是绝对有效或绝对正确的，理论才是研究的先导。当且仅当研究技术适于研究内容的特征和属性的时候，它们才能成为数据收集中强有力的工具，而且，要想对重要的问题进行系统阐释，那就只能在对该问题的普遍过程完全通晓之后进行，而不是在此之前。

表5.1,列出了上述三个场所中会用到的研究方法与研究技术,对这些常用的方法、技术,我们为读者提供了一些辅助性说明,而且,在这一部分的末尾,我们将提供一个参考书目表,这些书目能够详尽地说明各种不同方法与技术的有关细节。

表5.1 三种场所常用的研究方法与技术

场　所	方　法	技　术
图书馆	1. 历史记录分析:包括信件、日记等原始材料,二手的解释材料等 2. 文献分析:正式机构的数据或非数据记录等 3. 在各类书籍、刊物和专题论文中寻求理论或查找先前的研究	·便笺(note)记录 ·内容分析 ·磁带或胶片浏览或分析 ·数据编汇或处理 ·建立参考或摘要指南 ·检索微缩胶片或卡片 ·计算机信息咨询
实地	1. 邮寄问卷	·确定填答者的社会、经济背景 ·运用社会成员心理研究与测定量表判定如下变量:如社会地位,团体结构,社区、社会参与程度,领导行为和家庭调适等等 ·运用态度量表测量道德水平
	2. 个人访谈和结构式访谈(structured interivew)	·访问者运用开放性与封闭性问题,按详细计划进行访问 ·会用到社会成员心理研究与测定量表
	3. 焦点访谈(focused interview)	·访问者集中于对象的某段特殊经历及其产生的影响,预先应推测对象会掩饰哪些话题或问题
	4. 自由访谈	·让回答者轻松自如地谈论所研究的主题
	5. 团体访问	·同时访问由被访者组成的小规模的团体,上述技术都可以采用
	6. 电话调查	·可以作为一种获取信息、表达观点的调查技术 ·可以作为邮寄问卷的后继方法以提高回收率
	7. 案例与生活史研究	·对于案例研究,应收集有关个人的剖面资料以便于深入分析,强调社会化过程中的个人与社会因素 ·对于生活史研究,应收集有关个人的纵向资料以深刻认识其个性特征,同时也强调一定时段内发生的社会化过程
	8. 非参与性直接观察	·运用标准计分卡及行为观测量表
	9. 参与观察	·互动记录,可能的话,运用磁带录音或运用摄像技术
	10. 大众观察	·通过在公众场所独立观察或会谈,记录大众行为或集体行为
实验室	针对个人或团体的小群体研究:其偶发行动、游戏、问题解决或紧张表现等,亦可进行组织的、角色的分析	·运用创设的、非结构化的情境,或组织同盟,或应用视听记录装置,或立于单面镜后观察

表 5.2 研究技术的选择

问 题	途 径	研究技术
需在控制条件下获得可信的信息	在实验室测量对象	实验室内做实验,或模仿
想知道人们在公众场所的表现	观察	自然状态下观察
想知道人们私下的行为	说服对象记日记	个人文献研究
研究人们的想法	询问对象	会谈、问卷、态度量表
想知道人们的行踪	绘制行动图	行动路线图,行踪测绘表
需确定个性特征或评估心智能力	进行标准化的测试	心理测验
需鉴别言语中的倾向性	系统化的表格	内容分析
需理解一件不寻常的事情	详尽而较为漫长的调查	案例研究

资料来源:Robert Sommers & Barbara B Sommers:*A practical Guide to Behavioral Research*, 2nd *ed.* (Oxford university Press),1986,p.8.

Robert 和 Barbara(1986)在表 5.2 中,针对研究方法的选择提炼出一些有用而简明的规则,注意,他们是在绘出一份需研究问题的简要目录之后,再讲解适合该问题的研究手段及研究技术。他们指出:

> 观察法尤其适合研究公共场所的人们。对于私下行为,个人日记更有利于研究,而如果对于同一现象出现了不同的解释,那么实验法在区分这些解释时会成为强有力的工具。同样,要研究自然状态下人们的观念和行为,那就与其用实验法,还不如用问卷调查或访谈更为有效。

尽管上述规则非常有用,还是应该牢记一点,那就是:寻找某个问题答案的特定研究方法,总是极大地取决于该问题的提问方式,以及该问题所关注的核心内容。

参考文献

Sherif, Muzafer. (1967). In Gordon J. DiRenzo (Ed.), *Concepts, theory, and explanation in the social sciences.* New York: Random House.

Sommers, Robert, & Sommers, Barbara B. (1986). *A practical guide to behavioral research* (2nd ed.). New York: Oxford University Press.

5.2 浏览纸质或在线的社会科学研究资源

一般说来,能推进某种研究走向深层次的问题总是有各种不同来源的,而更经常的是,它们总起源于一些历史课题。若果真如此,那么搜寻原有的科学著述就显得极其重要。一些社会科学家常常忽略了这一步,只是因为他们往往会误认为自己的研究课题非常新颖,前人很少涉足,有太多的时候,他们不能对这一点给予足够重视:即他们的研究是建立在过去前人研究的基础上的,这些新研究可以证实前人的某种假设或理论,为前人的知识增添新的内容。总之,他们低估了前人工作的价值,也对其学科将来发展所提出的要求认识不足。

物理学家和生物学家大多不愿冒这种学术风险,因为他们知道,后来的学者会很快地觉察出他们工作中的这类失误。社会科学家由此应该意识到:其实他们也绕不开这一问题。

能用来引导和搜索学术研究著作的资源既十分丰富,又卓有成效。它们可以分成四类:期刊杂志(论文)检索、计算机辅助参考服务系统、微缩胶片—微缩平片—微缩印刷品媒介,以及特别制作的资料读物。

下述索引对于社会科学家价值颇大,它们是按查询时使用的频率编排的,依次是:社会科学索引、社会科学引文索引、当代教育期刊索引、人口索引。

5.2.1 社会科学索引(SSI)

社会科学索引是一种在数量上不断累积的、针对英文期刊或杂志的索引,其主体由期刊或杂志中的作者姓名、论文主题条目组成,其研究领域涉及:瘾癖行为研究、人类学、地域研究、社区保健与药物治疗、惩戒行为、司法公正、犯罪学、经济学、环境研究、伦理学、家庭研究、性研究、地理学、老年学、国际关系、法律、少数民族研究、公共计划与管理、政策科学、政治科学、精神病学、心理学、公共福利、社会工作、社会学以及城市研究。另外,在索引主体后面,还附有一个引用书目的作者姓名一览表。该索引由 H. W. Wilson 负责发行,拥有超过 607 000 份内容摘要,每年还在以多于 47 000 份的速度递增,它有印刷版本,也有电子版本,印刷版可以从 www. hwwilson. com 网站上购买。

社会科学索引的在线版(或 CD 版)可以从 www. sliverplatter. com 网站上得到,这个索引每年按季度发行,并在每次发行时必然添加一些新内容。它有相当长的历史,从 1907 年开始,它就以"国际索引"的名字出现,到 1955 年,它继续发行并更名为"社会科学与人文科学索引",但在 1974 年的时候,知识体系的急剧膨胀使得它应时而变,分成了"社会科学索引"和"人文科学索引"两种索引,这种划分一直持续到现在。

5.2.2 社会科学引文索引(SSCI)

社会科学引文索引始于 1956 年,持续到现在,按年、月、日编排,它为人们搜寻过

去或当今的目录学信息、作者摘要和参考资料提供便利,涵盖了 1 700 多种社会科学专业期刊,涉及 50 多门学科,还拥有 5 700 多个来自科学与技术杂志的条目,它可以从不同的途径或格式获取,比如通过 ISI Web 科学网站,能利用因特网或局域网得到,该网站从 1956 年起每周更新一次,直到现在。比如在 1992 年前,可以获得每月更新的、包括作者摘要的 CD 光盘,另外,通过 Social SciSearch 网站,可以在线获得从 1972 年起直到现在每周更新的资料。

对于这个索引来说,有价值的或许就在于它的"引用参考"搜寻功能,它可以让你查找并引用一些特定资源上的文章。例如,倘若你知道一个特定姓名的作者,你就可以搜索这个名字,然后看看他有什么出版物被收集在内,这种方式,对于想要建立某个学者作品目录的人来说,应该是很完美的。

社会科学引文索引既有印刷版,也有不同形式的电子版,由美国费城 3501 市场街 ISI 出版,电话(800)336-4474 或 215-386-0010,E-mail 地址:sales@ isinet. com。

这一索引涵盖了十分广阔的知识范围,具体包括:人类学、地域研究、商务、交际、犯罪学和刑罚学、人口统计学、经济学、老年医学和老年学、健康政策与服务、历史学与科学哲学、社会科学历史、工业关系与劳工、信息科学与图书馆学、国际关系、语言学、法律、管理学、药物学、护理学、哲学、计划与发展、政治科学、精神病学、心理学(应用生物心理、临床心理、心理发展、教育心理、实验心理、心理数学、心理分析及社会心理)、公共管理、公共环境与职业恢复、社会问题、社会科学(生物医学)、社会科学(各学科间的交叉研究)、社会科学(数学、方法学)、社会工作、社会学、特殊教育、有瘾品滥用、交通、城市研究以及妇女研究等。

"社会科学引文索引"操作起来相对简单,在搜索时,搜索的主题往往由参考的起点,而不是某个词或某个标题来确定,因而,搜索基本与某个特别的术语或某种机器语言没有关联。通常是,查询者由某个参考资料或某位作者开始进行查询(他们可以由某书的脚注,或某本书,或某种专业性学科全书,甚至是一些职业用语或标题来获得上述信息),接着,他便可以搜索该作者的姓名,在输入其姓名之后,查询者就能判定这位作者有哪些作品最值得参考。于是,电脑会显示该参考物,通过判断是哪年,哪种期刊,哪一卷,哪个页码,查询者由此还知道是谁刚刚引用过这些著作。在记录下所有引用过该作品的作者名单并汇编成目录之后,查询者就可以转到资源索引窗口并得到与该作品有关的全部的目录学数据。

一位查询者通过这种索引,能得到较快回答的基本问题是:"在哪里,由谁把该文章引用进去?",同时,它还被许多科学家利用,以确定该著作是否已被别人引用或批评过,当然,这样在科学的交流圈里能让人得到更多的反馈信息。其实,任何人都可以做这样的选择:忽略自己的作品如何被引用这一问题,而关注一些里边出现的、为其他社会科学家引介的内容。同时,"社会科学引文索引"也可以用来查询目前到底有哪些人在从事这个特定问题的研究,或者确定某个作品是否已经被引文,是否存在某个主题综述,甚至,也可以查询一个概念是否已得到应用,某种理论是否得到肯定,某种方法是否已得到提高。由于该索引上还会公布一些指导性的修正建议,所以它对于后来相关文章的撰写也能提供有益的帮助。总之,这种索引对于科学家和图书馆职员来

说,有说不完的好处,使用者想像力越强,应用它时也才会越少出现局限感。

一般而言,用户能从该索引上得到回答的问题有:

- 这个论文被引用过吗?
- 有关于这个主题的评论吗?
- 该理论是否已得到肯定?
- 这项工作还在扩展吗?
- 这种方法是否已被改进?
- 该建议是否被尝试?
- 这种理念果真是初创的吗?
- 这篇"将要出版的论文"现在出版了吗? 哪里可以找到它?
- 与这次初步交流有关的全文在哪里可以找到?
- 这种技术的相关报道是否已在某个期刊上发表?
- 是否于随后发表过相关的勘误表或更正声明?
- 这篇论文简介的数据在哪里?
- 在哪里能找到关于这种主题的评论文章的原始数据?
- 要改进这本书中这一章,是否有足够的新信息来确保?
- 什么样的原始资料才能用来分析网络上出现的原有统计图?
- 还有谁在从事该学术领域的工作?
- 有用来描述该研究领域状况的数据吗?
- 这种知识的潜在应用领域在哪里?
- 这种理论或观念是否已应用到一个新的领域?
- 什么样的出版物是发端于这一机构?
- 这篇论文是否已被摘录进基本期刊?
- 这位学者以第一作者名义发表的当前作品有哪些?
- 这位学者以第二作者名义发表的当前作品有哪些?
- 这位学者还有什么其他著作?
- 这位学者的作品是否已被汇编成作品集?

5.2.3 当前教育期刊索引(CIJE)

"当前教育期刊索引"上的期刊资源数据库(见网站:www. oryxpress. com/cije. htm)由 Oryx 出版社发行,并且每月给出当前教育期刊上的文献指南,它涵盖了大约980 种主要教育期刊或与教育相关的期刊上发表的全部文章,该索引是由教育资源信息中心(ERIC)发起的,由美国教育部、国家教育资料馆、教育研究与发展(OERI)等机构共同管理。其地址为:400 Maryland Ave,Sw,Washington,DC 20202;电话:(800)424-1616。

5.2.4 人口索引

人口索引（见网址：http://popindex.princeton.edu）从1935年开始建立，并成为研究世界人口问题的基本参考工具。它推出包含注释的资料目录，这些资料包括与“人口”这一主题有关的新近出版的图书，期刊上的文章、论文以及其他材料，它还为人口学家设立人口学研究的著作目录，其在线版本收集了从1986年到1999年发表于人口索引上的46 035篇人口学著作的摘要。对于这些著作，人们是再也见不到其印刷版了。

除了大约400种期刊作为固定来源之外，它还收集了其他期刊上的文章提供给人们参考，其内容有生物学、地理学、经济学以及社会学，分别归入到各学科的主要目录之下，查阅时大致可见到：生物学摘要、当前出版物（社会科学与行为科学）、当前地理学出版物、地理学摘要（人文地理学）、经济学著作（期刊）、国际PAIS出版物，以及社会学摘要。

所有目录都是按文章主题进行编排的，文章不重复引入，并且按照指南（在每个条目的批注上）进行分类，而在一定的主标题下面，引文则是按第一作者姓名的字母顺序进行编排的。如果要交叉检阅（同时查阅多个文章），那就得依赖主标题，一般来说，在每个条目的主标题下面，都有相关文章的编排号，人们会读到这样的句子：“亦可查……标题。”另外，如果可行的话，一些地理学与作者姓名的索引还会依照某个研究问题进行编汇、发布，这种汇编集每一册都有新内容添加进来，而且，波士顿的G. K. Hall和Co还把1935—1968年，以及1969—1981年的这些集子印刷发行过。

5.2.5 社会学摘要

社会学摘要由剑桥科学论文摘要（www.csa.com/）发行，有印刷版和在线版，联系地址为：7200 Wisconsin Avenue，Suite 601，Bethesda，MD 20814，(301)961-6700，其CD产品可以由Silver Platter（www.silverplatter.com/）处得到。它按月发行，每月大约能增加2 500种新内容，到1999年6月，其总记录已超过500 000。

社会学摘要录用20种主要信息领域（1963年至今）上发表的研究文章，其内容则涉及人类学、商务、集体行为、社区发展、灾害研究、教育、环境研究、性研究、老年学、法律与刑罚学、药物（或有瘾品）滥用、城市研究以及暴力行为。

对于每一个摘要，都可以通过下述体系代码获得有用信息。（表5.3）

表 5.3　社会学研究的相关检索代码表

体系代码	含　义
AB	摘要(Abstract)
IS	国际标准期刊编号(ISSN)
AF	作者简历(Author Affiliation)
LA	语言类型(Language)
AN	图书馆新添书籍编号(Accession Number)
LC	图书馆的国会管理号码(Library of Congress Control Number)
AU	作者(Author)
NT	备注(Notes)
CD	分类编号(CODEN)
TO	原始标题(Original Title)
CL	类别划分(Classification)
PB	发行者(Publisher)
CP	发行国家(Country of Publication)
PT	出版物类别(Publication Type)
DE	描述符,解说符(Description)
PY	发行时间(Publication Year)
IB	国际标准图书编号(ISBN)
SO	来源(Source)
ID	标识符(Identifiers)
IT	标题(Title)

社会学摘要对于社会科学问题研究中的学科交叉性研究来说,以及对于在多种学科中寻找社会学发展机遇的从业者来说,应该是基本的参考物,它既收录了书本,某个章节,有着关联的某些论文摘要,又设置了专业书、影视材料和应用软件,以供引用时检阅。

如果在社会学摘要上的社会学参考资料栏里能看到相关的论文复印件,那它就必定会是在每个摘要的备注(Notes)里,查阅者可依此搜寻。若想得到更多信息,可与该机构联系,方式是:LLC at P. O. Box22206,San Diego,CA 92192-0206,(800)752-3945,E-mail:info@ mail. socabs. com。社会学摘要往往也建立有主标题索引和作者姓名索引。这种每册都有新内容添加的索引,在每一年作为最后一期(第 8 期)出版,包含一份内容表、一份标题索引、一份期刊索引、一份著作索引、一份作者姓名索引以及一份缩写说明书。

从 1974 年开始,社会学摘要机构发行了社会学档案(CD 版),它收录了来自社会

学或相关学科的 1 800 多种期刊上的论文摘要或引用文章,并且从 1986 年至今,作为对这种引用目录的改进和提高,它加入了相关专题论文这一项目。另外要说的是,从 1980 年起,它还建立起社会政策规划与发展摘要资料库,里边有比较详细的文章摘要。

5.2.6 心理学摘要

心理学摘要(www. apa. org/psycinfo/products/psycabs. html)由美国心理协会发行,其作用与上述社会学摘要大致相同,涉及社会科学中的很多领域。

社会科学家常常从事相近领域的研究,所以,社会学研究者其实就像对待自己专业那样,也想从相关学科的文章摘要里边获得尽可能多的信息。然而,由于各学科提供的摘要服务总是力求独立操作的(就像人们常见的努力争取某种学科独立性那样),它们在总体上看来,会在内容上有所重叠,这种现象也客观地存在于每种学科内部。心理学摘要按月发行,建立了内容总结(包括摘要、目录信息、索引等),这些总结主要针对一些英文版期刊上的文章,某些技术报告,某些书本章节或者是心理学的某些书籍。它按如下研究内容把引文分成 21 个主要类别(还有更多的子类别),在这种结构中,每种类别的结构代码如下:

- 2100 普通心理学(General Psychology)
- 2140 历史与体系(心理学)(History & Systems)
- 2200 心理测验·统计学·方法论(Psychometrics & Statistics & Methodology)
  - 2220 测试与测验(Tests & Testing)
  - 2221 感官与运动知觉测验(Sensory & Motor Testing)
  - 2222 发展量表与规划(Developmental Scales & Schedules)
  - 2223 个性量表与细目表(Personality Scales & Inventories)
  - 2224 临床心理学测验(Clinical Psychological Testing)
  - 2225 神经心理学诊断(Neuropsychological Assessment)
  - 2226 健康心理学测验(Health Psychology Testing)
  - 2227 教育评估(Educational Measurement)
  - 2228 职业与就业测验(Occupational & Employment Testing)
  - 2229 消费意向与态度测验(Consumer Opinion & Attitude Testing)
  - 2240 统计学与数学(Statistics & Mathematics)
  - 2260 研究方法与实验设计(Research Methods & Experimental Design)
- 2300 人类实验心理学(Human Experimental Psychology)
  - 2320 感觉(Sensory Perception)
  - 2323 视觉(Visual Perception)
  - 2326 听觉与语言能力(Auditory & Speech Perception)
  - 2330 运动过程(Motor Processes)
  - 2340 认知过程(Cognitive Processes)

2343 学习与记忆(Learning & Memory)
2346 注意过程(Attention)
2360 心境与动机(Motivation & Emotion)
2380 良知状态(Consciousness States)
2390 灵学(Parapsychology)
2400 动物实验与比较心理学(Animal Experimental & Comparative Psychology)
2420 学习与动机(Learning & Motivation)
2440 社会行为与本能行为(Social & Instinctive Behavior)
2500 生理心理学与神经科学(Physiological Psychology & Neuroscience)
2510 遗传学(Genetics)
2520 神经心理学与神经学(Neuropsychology & Neurology)
2530 电生理学(Electrophysiology)
2540 生理过程(Physiological Processes)
2560 心理生理学(Psychophysiology)
2580 精神药理学(Psychopharmacology)
2600 心理学与人类本性(Psychology & the Humanities)
2610 文学与艺术(Literature & Fine Arts)
2630 哲学(Philosophy)
2700 沟通系统(Communication Systems)
2720 语种、语言与话语(Linguistics & Language & Speech)
2750 大众传播媒介(Mass Media Communications)
2800 发展心理学(Developmental Psychology)
2820 认知与知觉发展(Cognitive & Perceptual Development)
2840 社会心理与个性发展(Psychosocial & Personality Development)
2860 老年学(Gerontology)
2900 社会过程与社会问题(Social Processes & Social Issues)
2910 社会结构与社会组织(Social Structure & Organization)
2920 宗教(Religion)
2930 文化与人种学(Culture & Ethnology)
2950 婚姻与家庭(Marriage & Family)
2953 离婚与再婚(Divorce & Remarriage)
2956 小孩抚育与照顾(Childrearing & Child Care)
2960 政治过程与政治问题(Political Processes & Political Issues)
2970 性角色与妇女问题(Sex Roles & Women's Issues)
2980 性行为与性适应(Sexual Behavior & Sexual Orientation)
2990 药物与酒精使用(合法)(Drug & Alcohol Usage (Legal))
3000 社会心理学(Social Psychology)
3020 组织与人际交往过程(Group & Interpersonal Processes)

3040 社会知觉与社会认知(Social Perception & Cognition)

3100 个性心理学(Personality Psychology)

3120 个人特性与过程(Personality Traits & Processes)

3140 个性理论(Personality Theory)

3143 心理分析理论(Psychoanalytic Theory)

3200 心理失调与生理失调(Psychological & Physical Disorders)

3210 心理失调(Psychological Disorders)

3211 情感失调(Affective Disorders)

3213 精神分裂及精神病状态(Schizophrenia & Psychotic States)

3215 神经官能症与焦虑性失调(Neuroses & Anxiety Disorders)

3217 个性失调(Personaliy Disorders)

3230 行为失调与反社会行为(Behavior Disorders & Antisocial Behavior)

3233 物品滥用与瘾癖(Substance Abuse & Addiction)

3236 犯罪行为与青少年犯罪(Criminal Behavior & Juvenile Delinquency)

3250 发展性失调与孤独症(Developmental Disorders & Autism)

3253 学习失调(Learning Disorders)

3256 智力障碍(Mental Retardation)

3260 进食功能障碍(Eating Disorders)

3270 话语与语言失调(Speech & Language Disorders)

3280 有毒环境与保健(Environmental Toxins & Health)

3290 生理、身体结构与基因性心理失调(Physical & Somatoform & Psychogenic Disorders)

3291 免疫性功能失调(Immunological Disorders)

3293 癌症(Cancer)

3295 心血管功能失调(Cardiovascular Disorders)

3297 神经性失调与大脑损丧(Neurological Disorders & Brain Damage)

3299 视觉、听觉与神经传导功能失调(Vision & Hearing & Sensory Disorders)

3300 健康、精神健康预防及治疗(Health & Mental Health Treatment & Prevention)

3310 精神疗法与心理咨询(Psychotherapy & Psychotherapeutic Cunseling)

3311 认知疗法(Cognitive Therapy)

3312 行为疗法与行为矫正(Behavior Therapy & Behavior Modification)

3313 团体与家庭疗法(Group & Family Therapy)

3314 互动治疗,病人中心主义疗法及人本主义疗法(Interpersonal & Client Centered & Humanistic Therapy)

3315 心理分析疗法(Psychoanalytic Therapy)

3340 临床精神药理学(Clinical Psychopharmacology)

3350 特别参与(干涉)(Specialized Interventions)

3351 临床催眠(Clinical Hypnosis)

3353 自助团体(Self-Help Groups)
3355 非专业性咨询或半职业性咨询或牧师咨询(Lay & Paraprofessional & Pastoral Counseling)
3357 艺术疗法、音乐疗法及运动疗法(Art & Music & Movement Therapy)
3360 心理保健与药物(Health Psychology & Medicine)
3361 行为疗法或生理疾病的心理治疗(Behavioral & Psychological Treatment of Physical Illness)
3363 生理疾病的药物处理(Medical Treatment of Physical Illness)
3365 保健、健康改善及健康维持(Promotion & Maintenance of Health & Wellness)
3370 健康与心理健康服务(Health & Mental Health Services)
3371 院外病人服务(Outpatient Services)
3373 社区及社会服务(Community & Social Services)
3375 家庭照料与收容所(Home Care & Hospice)
3377 家庭护理与邻里照料(Nursing Homes & Residential Care)
3379 住院病人与医院服务(Inpatient & Hospital Services)
3380 健康恢复(Rehabilitation)
3383 致瘾物与酒精等的摆脱(Drug & Alcohol Rehabilitation)
3384 职业与再就业(Occupational & Vocational Rehabilitation)
3385 言语与语言疗治(Speech & Language Therapy)
3386 罪犯教育学与重新做人(Criminal Rehabilitation & Penology)
3400 职业心理与健康问题(Professional Psychological & Health Personnel Issues)
3410 职业教育与培训(Professional Education & Training)
3430 职业态度与职业特征(Professional Personnel Attitudes & Characteristics)
3450 职业道德、标准与义务(Professional Ethics & Standards & Liability)
3470 职业伤害(Impaired Professionals)
3500 教育心理学(Educational Psychology)
3510 教育管理与受教育(Educational Administration & Personnel)
3530 课程、规划与教学方法(Curriculum & Programs & Teaching Methods)
3550 学术研究与成就(Academic Learning & Achievement)
3560 学习动力、学生调控与学习态度(Classroom Dynamics & Student Adjustment & Attitudes)
3570 特殊教育与补习(Special & Remedial Education)
3575 天赋与潜能(Gifted & Talented)
3580 教育咨询、职业咨询与学生服务(Educational/Vocational Counseling & Student Services)
3600 工业心理学与组织心理学(Industrial & Organizational Psychology)
3610 职业兴趣与职业指导(Occupational Interests & Guidance)

3620 雇员挑选、管理与培训(Personnel Management & Selection & Training)
3630 员工评估与职业表现(Personnel Evaluation & Job Performance)
3640 管理与管理技能培训(Management & Management Training)
3650 员工态度与职业满足感(Personnel Attitudes & Job Satisfaction)
3660 组织化行为(Organizational Behavior)
3670 工作环境与工业安全(Working Conditions & Industrial Safety)
3700 运动心理学与休闲(Sport Psychology & Leisure)
3720 运动(Sports)
3740 休闲与娱乐(Recreation & Leisure)
3800 军事心理学(Military Psychology)
3900 消费心理学(Consumer Psychology)
3920 顾客态度与行为(Consumer Attitudes & Behavior)
3940 市场与广告(Marketing & Advertising)
4000 基因工程与环境心理学(Engineering & Environmental Psychology)
4010 人类基因工程(Human Factors Engineering)
4030 生活空间与机构设计(Lifespace & Institutional Design)
4050 社区与环境规划(Community & Environmental Planning)
4070 环境问题与态度(Environmental Issues & Attitudes)
4090 交通(Transportation)
4100 智能系统(Intelligent Systems)
4120 人工智能系统与专家(Artificial Intelligence & Expert Systems)
4140 机器人(Robotics)
4160 神经网络(Neural Networks)
4200 法医心理学与法律问题(Forensic Psychology & Legal Issues)
4210 公民权利与公民法(Civil Rights & Civil Law)
4230 刑法与刑事判决(Criminal Law & Criminal Adjudication)
4250 调解与解决冲突(Mediation & Conflict Resolution)
4270 犯罪预防(Crime Prevention)
4290 警察与公民守法(Police & Legal Personnel)

5.2.7 心理学信息

心理学信息(www. apa. org/psycinfo/about/)是心理学摘要的电子版本,涵盖了1 646种期刊,涉及与心理学摘要的"通信版本"(印刷品)相同的主要领域,里边的记录超过了1 000 000条。

5.2.8 联邦政府教育资源

联邦政府在网站(www. ed. gov/Edres/Edfed/)上拥有下述电子资源,可提供教育

及相关领域的参考信息。

（1）地区综合辅助中心

地区综合辅助中心（www. ed. gov/EdRes/EdFed/EdTechCtrs. html）由15个分支综合辅助中心组成，作为整个网络机构的组成部分，它们遍布全国，为人们提供帮助或信息。他们在该网络中担当的角色，就是帮助地方政府、学区，或者满足由基本与初等教育法案（ESEA）规定的在校孩子们的需求。这种需求的对象有：极端贫困地区的孩子、流浪孩子、移民孩子、英语能力有限的孩子、受忽视或逃学的孩子、无家可归的或年幼的印第安孩子、残疾孩子，并且，在适当情况下，还包括阿拉斯加州与夏威夷州的本地孩子（参见法案第八条）。不过，辅助中心重点关注法案第一条所规定的教育项目，积极协助地方教育机构。

（2）教育资源信息中心（ERIC）

教育资源信息中心（www. ed. gov/Edres/EdFed/Eric. html）是一种全（美）国的信息网络，它检索资料，编定目录，作出归纳，为人们获取不同资源上的教育信息提供便利，它所收集的数据与资料库被世界3 000个左右的地方所收藏，其中包括绝大多数主流的公共图书系统和大学图书系统。该中心拥有相当多的出版物，并提供广泛的用户支持，“中心访问”就是其中一种，它为在线教师提供网上咨询服务，另外，“中心访问虚拟资料馆”（见网址 http://askeric. org/）还是极富个性化的、基于因特网的服务机构，它为整个美国的教师、图书管理员、咨询师、管理者、父母亲以及其他所有对教育感兴趣的人提供教育信息，这些信息包括课程的教学计划集，如何利用教育资源信息中心、邮件联系地址列表，以及“中心指南”等。

附挂在“教育信息中心”系统上的机构有“早期儿童研究与实践”（www. ecrp. uiuc. edu/），它是关于幼童发展、照料与教育的网上杂志。另外，就是“美国父母所需信息网”（http://npin. org/），它为父母们或帮助他们的那些机构提供信息与交流平台。

教育资源信息中心系统上有16种特定主题的信息（情报）交换所，有“中心”参考与使用工具，有访问系统，该访问系统为社会科学家或行为科学家提供介绍性服务，并提供其他有用资源。对于“中心”来说，最有利用价值的可能是其主要的以及分支的信息（情报）交换所，每个信息（情报）交换所都会通过收集、编选摘要，建立索引等手段，为“中心”的数据资料库提供教育方面的材料，它们也会在各自范围内回答特定主题的信息请求，或者发行与当前研究、规划及实践相关的特定出版物。实际上，像这些交换所中的任何一个，都有十分巨大的信息量，它们的电子域名及内容显示如下。不过，值得注意的是，这些交换所会因为各自关注的主体不同而产生不同的格式，拥有不同的收集材料。尽管如此，它们都可以通过上面提及的“中心”主网站进入。

- 成人、职业生涯与职业教育（http://ericacve. org/），包括各种不同水平、不同方面的成人教育、继续教育、职业生涯培训以及职业或技能培训。
- 评价与评估（http://ericae. net/），提供特别的教育评估与测试手段。
- 社区学院（www. gseis. ucla. edu/ERIC/eric. html），主要针对两年制的公立或私立社区学院、初等学院、技术学院以及各种大学分院，为它们提供发展、管理与

评估方面的指导。

- 咨询与学生服务(http://ericcass.uncg.edu/),包括为各种不同教育水平,处于不同状况的咨询者提供预备、实习或指导等方面的信息,以及理论发展咨询和其他学生服务。
- 残障或天才教育(http://ericec.org/),涵盖了残障者或天才者各个方面的教育与发展,具体包括其鉴别、评判、干预、充实等,既照顾到特殊状况的,也包含了在主流教育机构内学习的。
- 教育管理(http://eric.uoregon.edu),包含各种层面的管理方法、领导艺术和控制手段,以及那些处于基本或初等水平的公立或私立教育机构的机构设置,还包括为其实际运作提供实物、设备等问题。
- 初等教育或早期儿童教育(http://ericeece.org/),包括从出生到青少年早期这种儿童的身体成长、认知发展、社会化、教育及文化的发展等内容。
- 高等教育(www.eriche.org/),发表关于学院或大学的问题、发展计划、学生状况、课程设置、教育计划,以及研究成果等。
- 信息与技术(http://ericir.syr.edu/ithome/),包括各种水平的教育技术、图书馆以及信息科学。
- 语言与语种(www.cal.org/ericcll),指语言及语言科学,包括各种不同层面的第二语言教育指导,以及所有普通或特殊教学语言的学习。
- 阅读、英语与沟通(www.indiana.edu/~eric_rec/),包括阅读、英语运用、口头或非口头的沟通,以及大学预备教育。
- 边远地区教育与小规模学校(www.ael.org/eric/),包括与边远地区居民教育计划与操作有关的各种经济、文化与社会环境条件,以及美国印第安人、阿拉斯加本地居民、美籍墨西哥人、其他移民等的教育,还包括所有小规模学校的教育规划与实践、户外教育等内容。
- 科学、数学与环境教育(www.ericse.org/),包括各种层次的科学、数学与环境教育。
- 社会研究/社会科学教育(www.indiana.edu/~ssdc/eric_chess.htm 或 http://ericso.indiana.edu),主要关注一些学科的教育与学习中会出现的问题,这些学科包括社会研究或社会科学中的历史、地理、公民学、经济学等等。
- 教学与教师培训(www.erisp.org/),包括教师招聘、挑选、文凭、证书、培训、人员维持、职业准备、评估、激励、退休安排等等,也涵盖了诸如保健、生理教育、休闲、跳舞等多方面内容。
- 城市教育(http://eric-web.tc.columbia.edu/),包括城市学校的规划与运作信息,美籍非洲青年和美籍西班牙青年的教育,教育公平的理论与实践,城市与小群体教育历程,社会机构以及其他服务。

教育资源信息中心上的附属信息(情报)交换所被安排在每一个主要交换所之下,并为这些特定主题的交换所提供与该主题相关的全部或部分功能。具体包括:为

他们辨识、获取符合该领域内容的重要著作，并收入“中心”的资料库，提供参考或推荐服务，给予技术帮助，维护或致力于网站建设，发行出版物等等，这些附属信息（情报）交换所从“中心”机构之外的发起者那里获取经费，它们包括如下内容：

- 小孩照料（http://nccic.org/），包括小孩照料相关联系（查询）信息体系——作为保障照料质量的一种机制，为小孩和其家庭提供的综合服务等。它由国家小孩照料信息中心所发起。
- 实验学校（www.aacte.org/ERIC/pro_dev_schools.htm），它提供有关实验学校、职业发展学校、合伙学校、职业练习学校以及相似机构的一些信息，它由美国学院教师教育联合会发起。
- 教育机遇获取（www.trioprograms.org/clearinghouse/），它提供各类信息，以帮助一些个人、父母或组织能有更多途径获得高质量的教育资源，这些对象主要是想了解如何或怎样才能让那些低收入者，（历史所称的）第一代人，以及残疾学生进入大学，它由美国教育机遇获取委员会发起。
- 企业家协会教育（www.celcee.edu/），它以建立索引、编汇摘要或传播的方式，提供企业家协会的教育信息，并促成这些信息资源为每个教育区域所享用，它由企业家协会领导中心发起。
- ESL（英语）文化教育（www.cal.org/ncle/），它为英语运用能力有限的成年人或非在校青年提供各方面的文化教育。它由国家ESL文化教育信息（情报）交换所发起。
- 国家法律（相关）教育资源中心（www.abanet.org/publiced/nlrc.html），作为辅助性交换所的一个，主要是收集与传播与法律教育有关的信息，包括教育计划、资源、实体法论题、法律基金资源、相关教师以及资源领导者的培训机会等等。
- 测试工具汇编（http://ericae.net/testcol.htm），它能提供有偿的以及无偿的测试试卷、核查表、仪器、问卷以及其他评估或判断工具。该测试域名（网络定位）实际上是由多家单位提供运行方案的，这些单位包括教育资源信息中心的评估与诊断信息（情报）交换所，图书馆与参考服务体系的教育测试服务分部，位于林肯郡 Nebraska 大学中的 Buros（智力）测量机构，华盛顿大学的第三区综合中心，以及赞成教育测试的发行商。
- 美国—日本研究（www.Indiana.edu/~japan），涵盖了与日本的社会、文化有关的教育和学习等各方面事务。

国家教育机构信息（情报）交换所（www.edfacilities.org），也附属于教育资源信息中心，它收集、管理和发布与教育机构相关的信息，作为一种资源，它也为在校人员，以及那些规划、建构与掌管教育机构的人员提供服务。

教育资源信息中心拥有资料复印服务部，可以制作、出售包含该中心资料摘要的微缩胶片与复印文本，它还提供“中心”过期文集，年度订阅安排，累加性索引和其他“中心”相关产品。

而 Oryx 出版社则出版当前教育期刊索引，教育资源信息中心解说汇编以及其他

一些与"中心"有关的产品。

美国政府印刷品办公室(www. access. gpo. gov/)则出版数额巨大的信息文件。

此外,还有艾森豪威尔国家信息(情报)交换所与地区协作会(www. mathscien-cenetwork. org/connect. htm),它可以为在数学和其他科学教育中出现困难的人提供技术援助或职业发展帮助。其十个地区信息(情报)交换所的每一个,都在为该地区提供重要的服务,并合而为一,给整个国家提供服务。

(3)公平帮扶中心

十个公平帮扶中心(www. ed. gov/EdRes/EdFed/epuity. html 上有查询指南)都由美国教育部提供赞助,并遵从 1964 年公民权法案中的第 IV 条规定。这些中心作为一种促进教育公平的手段,能为那些公共学区里的人们提供包括种族、性别初始平等权利等范围的帮助。

(4)国家研究与发展中心

国家研究与发展中心(www. ed. gov/offices/OERI/ResCtr. html),得到一些大学背景的国家教育研究与发展中心的支持,关注全国性的重要问题与议题。该中心拥有一些特定的主题,如早期儿童学习与成长,学生学业与成就,文化、语言的多样性与第二语的学习,后中学时期的成长,成年人教育以及教育政策,等等。

(5)地区教育实验室

全国共有十个地区教育实验室,每一个(www. ed. gov/EdRes/EdRed/RegLab. html 上有查询指南)都致力于帮助教育者、(教育)政策制订者解决他们所在州或地区存在的教育问题。这些实验室研究教育难题,印刷出版物,并为教师和教育管理者提供培训计划,每个实验室都发行了其出版物目录,这些目录涉及较大范围的话题,比如教学策略、学校提升以及父母参与等。

(6)地区教育技术协会(R*TEC)

按照地区教育技术协会(www. rtec. org/)的规划,要对各个州或地区的教育机构、教师,学校图书馆,使用传媒的某个单位,教育管理者以及其他教育实体提供帮助,使得一些教育技术能成功地结合到幼儿园(通过其 12 个级别)教学中,结合到图书馆传媒中心上,以及结合到像成年人文化中心这样的一些教育设置(机构)上。

(7)特殊教育与康复服务基金

残疾人辅助技术服务与系统(基金)改造工程(ATFSCP)(www. ed. gov/EDRes/EdFed/specedrs. html)为全(美)国性的辅助技术基金问题以及系统改变问题提供人员培训,技术帮助和有关信息。该工程目的就在于提倡人们获取更多的必要知识与技能,以改善残疾人辅助技术设备,为他们提供更好的服务。

(8)"星级学校计划"站点

"星级学校计划"(www. ed. gov/EDRes/EdFed/star. html)鼓励人们在数学、其他科学和外语学习上作出更有效的指导,它也支持人们提高艺术技能,加强职业教育。通过运用网络这种电子交流媒体,它的目标主要在于:帮助一些不能充分享受社会服务的人,包括那些文盲、处于不利地位或英语能力有限的人,还有残疾者。

美国国家教育部的这类计划可以在该部的网站上审阅或查询,其网址为 http://search. ed. gov/csi。

5.2.9 儿童成长(论文)摘要及目录汇编

从 1927 年到 2001 年,儿童成长(论文)摘要及目录汇编(CDAB)(www. srcd. org)一直由儿童发展研究协会主办,主要是摘要汇编,内容则包括社会学、心理学、小儿科学、社会工作、人类学、保育学、教育学等不同学科的三百多种期刊。

这些摘要被分类编进以下几个学术领域:生物学、个性与社会心理学、认知发展、精神病学、理论与方法论、教育学。另外,这些被引用的期刊还有书本介绍词条。

5.2.10 人口在线(POPLINE)

"人口在线"(www. jhuccp. org/popline/)是世界最大的、有关人口家庭计划及人口健康等问题的百科全书式数据资料库,它拥有 275 000 种以上的摘引记录,由 Johns Hophins 卫生与健康学校的"人口信息计划"处主持。"人口在线"保持每月更新,其 CD 版每六个月更新一次,每次更新时可以添加大约 10 000 项记录,而它上面的人口信息栏目,则每两周更新一次。

5.3 美国人口普查署及其他联邦(人口)信息资源

应该说,十年一度的美国人口普查(www. census. gov)是美国付出的最大努力之一,对于研究者来说,也是可以获得的最丰富的信息资源之一。而现在,它的资料在网上也可以得到了。从目前来看,要检索人口信息,最好从美国的 FactFinder(事实搜集者)网站(http://factfinder. census. gov/servlet/BasicFactsServlet)上开始,它是通向庞大的、与人口有关信息的入口,提供了方便地查询参考资料、数据资料、(人口)地图等内容的方法。

通过美国的当前人口调查、家庭调查以及收入与计划参与调查等项目,人们可以从"当前人口报告"网站(www. census. gov/population/www 或 www. census. gov/hhes/www)上获得这些十年一度的人口普查更新资料,这些资料包含如下信息:

- 婚姻地位(marital status)
- 家务与家庭(households and families)
- 校园新生(school enrollment)
- 教育参与(educational attainment)
- 生育(fertility)
- 年龄(age)
- 计算机使用(computer use)

- 家庭经济(family economics)
- 儿童照顾(child care arrangement)
- 健康照料(health care)
- 人口预测与人口工程(population estimates and projections)
- 收入与贫困(income and poverty)

另外,"人口状况"(www. census. gov/apsd/www/censtats, html)目前可以免费浏览、使用,它也是从美国人口普查署获取人口资料的一个工具,从它上面能得到成百上千的数据库资料。

问卷调查:设计与方法 5.4

邮寄问卷由一系列用来收集某些信息或观点的问题组成,它直接邮寄给那些按一定方式挑选出的准填答者,要求他们完成问卷的填答之后,再将该问卷回邮给调查者。

这种收集信息的方式运用较为普遍,因为它在花费很少的时间与资金的情况下,还能较有把握地获得真实的资料。当然,这种方法要想完全盛行,也会遇到阻力,其中之一就是填答者往往会因为他们手中问卷的问题过多而产生较大压力。由于问卷填答会占用他们越来越宝贵的时间,填答者会在填答之前仔细审思:该问卷调查的研究目的是什么?合作者是谁?调查结果是否对他们有益?完成问卷填答大致会花多长时间?这份问卷质量如何,是否通俗易懂?甚至他们还会关注到问卷的纸质等问题。

因此,作为至关重要的衡量因素——回收率,就会随着情况的不同出现很大差异。表5.4是有关研究的简洁归纳,显示了如下一些情况,比如:被调查的对象(人口)、调查的目的与日期、问卷的长度与邮寄的数量、回收的问卷数量以及占邮出总数量的比例,以及跟踪访问过的人口数量,正如读者所见到的,这些问卷在回收率上出现很大差异。

表 5.4　样本举例:具有不同特征的问卷与它们的回收率

调查对象	问卷目的	问卷长度	邮寄发出数量	它们的回收数量	回收率(%)	跟踪访问过的数量	研究机构
俄亥俄州,Summit 县“家庭谱系社会”地方分会的成员			221	115	52		
派克(Park)学院,马里兰州大学的大一、大二学生	测量拥有不同感受的群体间关系以及整体满意度	100 个项目		566	60	邮件及电话	
田纳西州卢瑟福地区的专业健康照料人员	研究这些专业照料人员中存在的有瘾物滥用问题		1 817	775	43		
堪萨斯州约翰逊郡社区大学的毕业生,他们在早期的学习年代完成过职业教育课程	评定(教育)机构的成效			515	95		
俄亥俄州 21 个成人监狱里的每一位高级图书管理员	收集有关俄亥俄州监狱图书馆及其负责人员的资料	36 个项目	21	12	57		
田纳西州的农场主	研究目前农村地区的社会环境			531	75		
从田纳西州的“职业培训参与法案”计划中途退出的人员			2 462		29.4	两份邮件	
来自弗罗里达州大城市学区的教师	研究在教学职业中的摩擦		310		53		
台湾 15 个职业培训中心接受职业培训人员	研究这些人员的满意度是否与其职业经验、年龄、职业类型、培训方式、教育水平、工厂工作经验以及执教老师自身培训的内容等因素有关		649	420	64.7		

亚拉巴马州的州代表及州议员	研究该州的立法者对于那些发表与工作相关的信息的新闻媒体的评定，是如何有赖于他们对这些媒体作为反思者与舆论机构这一角色的认知状态		140	82	59		
刚考上硕士或博士的往届毕业生	研究这些毕业生在刚开始工作时，他们对于生理治疗学家这一职业角色的理解，以及自我如何适应这一角色的理解是否发生了变化，以及在这些理解与其职业满意度之间是否存在某种关联		15		93		
美国社会中的儿童虐待问题治疗者	研究普遍存在的、儿童性虐待问题治疗者的筋疲力尽状态（1995 年）	Maslach 的“精力耗尽状态”调查细目表，个人极限问卷表，加利福尼亚心理学非固定尺度测量清单以及人口统计学数据	600	206	40	无相关报告	Barton Tay，Trent ham 主持的论文研讨
女大学毕业生	调查女性的自我评价对自身的影响，以及她们在成就动机与职业现实之间存有的内在冲突	Hoffman 的细目表，Hazan 和 Shaver 的测量表，Rosenberg 和 Hampilo 的测量尺度，工作实践索引，有关成就大小核查的辅助表格，Sadd 的“畏惧成功”测量尺度，人口统计问卷和生活标准问卷	453		30	无相关报告	Daniella Dankne 主持的论文研讨

续表

调查对象	问卷目的	问卷长度	邮寄发出数量	它们的回收数量	回收率(%)	跟踪访问过的数量	研究机构
路易斯安那州的州法官	测量他们对于小孩照管和探望等问题的态度			190	31		
圣弗朗西斯科以及圣克鲁斯的居民	调查这些市民对发生在1989年的“Loma Prieta”地震灾难的反应			1 652	46		
新泽西州从1984年到1987年间离婚的父母亲	调查他们对法庭作出的有关小孩照管安排的指令是否满意			112	11		
威斯康星州的校内心理学家	调查他们对学校里有关“艾滋病”教育与防范宣传等类似服务行为的参与程度		150		61		
医生	调查他们身上存在的非反应性偏见,以及考察跟踪邮件的有效性		408		35		
美国东南部一个人数很多的大学里的学生	考察在四种反应级别的刺激物作用下所引发的单一效果或交互影响						

5.4.1 邮寄问卷法的优点与缺点

每一位选择邮寄问卷法做调查的研究者都应该考虑到,在现代这种激烈竞争的社会背景下,大部分被调查者会想到节省时间,结果他们可能在没有填答完全的情况下就把问卷回邮了。有鉴于此,研究者最好能详尽地考察这种方法的优点与缺点(下有说明)。我们把缺点放在前面讲解,实际上要突出它们的不可忽视性,只有在总体上优点超越了缺点,而使用方法刚好适应该研究的时候,我们才会说这种问卷是比较适合的,在列出了邮寄问卷法的优点与缺点之后,我们给出了一份问卷组建指南,以及一份提高问卷回收率的技术指南。

(1)邮寄问卷的缺点

正如 Wallace 在 1954 年所指出的,研究者们已发现邮寄问卷法存在下述缺点:

1)不作出回答性反应应受重视。

- 那些由私人或操作技术相对较差的人发出的邮寄问卷,其回答率往往不超过 50%。
- 要提高回答率,就要作出较大的后继努力。

2)由于存在不回答者,那些仅仅由回答者构成的实际样本有可能会很大地偏离预计样本,从而使调查发生失误。

- 不回答者最终会形成我们对其一无所知的对象群体。
- 尽管如此,我们还是应该作出特别努力,以评估一下不回答者与回答者之间到底存在哪些差异(努力方法可以是使用挂号信、打电话、个人会谈等)。
- 在后继努力中,最需要面对的往往是那些无法确知其地址,无法得到其地址,以及无法通过邮寄方式联系上的人。事实上,这些由"无反应性回答"(no response)或"拒绝回答"(refuse to answer)的人所组成的群体数量有时相当大。

(2)邮寄问卷法的优点

同上,正如 Wallace 在 1954 年所指出的,研究者们也认为邮寄问卷法存在下述优点:

- 它使得人们在花费极少的金钱与精力的情况下,能覆盖宽广的调查对象面。
- 它提供了更为广泛的地理领域上的接触对象。
- 它能联系上那些住址难以确定以及难于会见到的对象。
- 它提供了更大的调查面,通过获得更大、更有代表性的样本而更大程度上保证了调查的有效性。
- 它能获得更多的、数量客观的回答者。
- 在填答者必须先核查信息再做解答的情况下,这种方法更为有效。
- 在相互商议之后能提供更多有效信息的情况下,这种方法更为有效。
- 在问题的排列方面,这种方法更能保证其一致性。
- 给回答者一定的匿名性。
- 提供了一种较为简便的、随时间推移而不断得到新情况的方式。
- 它减少了调查者的压力。

参考文献

Wallace, David. (1954). A case for—and against—mail questionnaires. *Public Opinion Quarterly*, 18, 40-52.

5.4.2 问卷设计与制作指南

制作问卷要花费较多时间与精力,但只要能正确地设计和控制问卷,那它就能成为重要而精确的测量手段。以下是一个问题栏目,依次指出了在问卷设计与制作中必须考虑的一些重点(请参考 Lazarsfeld1972 年出版的书,尤其是第 8 章,以及 Dillman1978 年出版的书,以获取更多信息)。

1)重新审查调查方法与调查课题、研究假设之间的关系。对自己研究的领域要有整体的把握,对研究的目的、所需数据的属性要有清晰的认识。

在一份描述性调查中,研究者应收集与该课题有关的,尽可能广泛、精确的信息,在一份有关某种理论的解释性调查中,研究者应关注某个因变量(或事实)的决定因素并检验一些特定假设。而在上述任一情况下,节俭与高效都是必须保持的重要原则。我们的法则是:收集所要的数据,但绝不超越这一界限。所以,在事先,我们应该清楚,我们将会用到哪些数据来做分析。必要时,可以先勾画草图,以描述与研究课题或研究假设相关的数据分布状况,变量关系状况,并在后来按实际情况加以修正。

2)问题的标准化。在设计问卷时,下述各点应加以考虑:

- 问题的语言始终应符合填答者的知识水平,调查者只有在某些特殊的填答者面前才能使用他们所熟悉的专业术语,而在普通大众面前则必须使用日常用语。
- 尽量使用对每个人来说都具备相同含义的词汇,例如,在调查美国人和英国人时可能会出现这样的问题:"你喝茶多不多?",那么,对于这个问题中的"茶",美国人通常会理解为一种饮料,而英国人则更有可能认为它是一种午餐后的小吃。
- 避免问题过长。一旦问题过长,其意思就会变得模糊、混淆不清。
- 不要事先就假定那些填答者对你的问题有确切的答案,或明确的观念。一位母亲可能知道她的孩子在读些什么书,但我们只有通过问孩子,才能真正知道孩子们读这些书的感受。
- 在自己内心应建立一个判别优劣的标准。比如,要获得准确性与即时性,与其问"你看过多少本杂志?",就不如问"你现在在看什么杂志?"
- 在陈述一个问题时,要么给出所有可能情况,要么就干脆不提示任何情况。例如,与其问别人:"你认为当你丈夫在家的时候,他应该帮助小孩穿衣服并给他喂食吗?"就不如这样问"你认为,当你丈夫在家的时候,他应该帮助小孩穿衣服并给他喂食,还是认为在任何情况下这些都应该是妻子的责任?"也可以这样问:"如果丈夫与妻子都在,谁应该帮小孩穿衣、喂食?"
- 应避免伤害回答者的自尊心。比如,不要问"你知道最高法院里首席法官的名字吗?"而应该问"你听说过最高法院里首席法官的名字吗?"
- 如果我们问到某种不愉快的经历,就应该先给回答者一个时机,让他能进入某

种积极的心态,从而避免他们从不够好的角度看问题。例如,不要问别人"关于X,你不喜欢他什么?",相反,我们可以先问他:"对于X,你有什么看法?"

- 要明确这一点:我们是需要一个直接的问题,还是一个间接的问题,或者是一个直接问题之后再附一个间接问题。例如,我们可以问这样一个直接问题:"在上班时你曾偷过东西吗?",而如果我们想变成一个间接问题,就可以这样问:"你知道有谁在上班时偷过东西?"
- 预先确定好某个问题应该是开放式的(允许回答者给出多种答案)还是封闭式的(回答者只能挑选出一个或几个答案)。例如,作为一个开放式问题,我们可以问:"人们认为在这个社区,一些人拥有着太多的特权,你认为这是真的吗?这些人又是谁呢?",而如果转化为封闭式的问题,我们就应该这样问:"人们认为在这个社区,一些人拥有着太多的特权,你赞不赞成这种说法?"
- 预先确定,我们是需要一个普通问题还是一个特殊问题。有时候,可能像这样问就足够了:"你对这本书欣赏程度如何?",但有时候,像下面这样问可能就更完美一些:"你把这本书推荐给别人了吗?"
- 避免运用像"你喜欢你所做的事情吗?"这样一些模棱两可的语句,我们可以换一种问法,比如:"你喜欢写小说之外的东西吗?"
- 避免带偏见或带引导性的问题。比如,不应该问"作为一位美国公民,你在最后一次选举中投过票以实践自己的选举权吗?",而应该问"你在最后一次选举中投过票吗?"
- 以合适短语提问以避免产生不必要的不愉快感觉。比如,与其问"你是从高级学校毕业的吗?",还不如问"你所完成的学业中,最高的毕业学校是什么?"
- 鉴别一个问题是个人性的还是非个人性的,以求获得更好的效果。比如,若是问"你上班的地方工作环境是否令人满意?",那就不如问"你对你的工作环境条件是否满意?",后者可能得到更有利于调查的回答。
- 每个提问应该只限于一个主题,或一个参考前提。例如,我们不应该问"你是赞成还是反对日渐增强的工作安定感以及可靠的年薪?",而应该把这个问题分开来提问,"你是赞成还是反对日渐增强的工作安定感?",以及"你是赞成还是反对一份可靠的年薪收益?"

3)对问卷进行编排。所问问题的排列顺序是很重要的,在编排这些问题时,我们应该考虑以下几点:

- 以易于回答的问题作为开始,填答者才有兴趣作回答。
  a. 不要一开始就问有关年龄、职业或婚姻地位等问题。
  b. 开始时应问一些能激起填答者兴趣的问题。
- 不要让填答者感到,前边的问题变成填答后面问题的某种条件。
  a. 安排时,可以从一般到具体。
  b. 安排时,可以从简单过渡到复杂。
- 排列好问题顺序以减轻填答者的敏感性,可以把收入等涉及个人隐私的问题放到后面。

- 确定如果想得到理想的信息,是只须一个问题还是要问多个问题。
- 对于自由回答式的问题,有时候让这些问题成对出现更为有益,我们可以就一个问题的正、反两个方面来提问。
- 结尾的开放式问题,是需要填答者仔细思考和组织语言的,因而数量应尽可能的少,一般我们总是把这样的问题放在最后,主要是想确保填答者能够作出回答。
- 话题以及问题应该加以排列,以使填答者觉得合乎逻辑。这样做的目的其实为了让填答者在回答时既自然又轻松。

4)对问卷进行试用。当进行问卷试用时,能清楚发现需改进的错误之处,这在问卷质量控制中是基本的一个步骤。

- 在你预备调查的人中选出一部分填答者作为代表,并同他们会面,在他们浏览你的问卷之后,鼓励他们提出任何与问卷内容相关的问题,同时,检查那些可能引起误解,模棱两可或导致他们自我防御的地方,征询他们的意见,让他们试着用另外的方式表达那些难以理解或难以回答的问题。
- 绝对不可省略"试用"这一过程。
- 慎重挑选纸张与字体,选择好的字体能在好的纸张上取得更好的效果,即便是油印的问卷看起来也能像打印的。
- 考虑怎样才能激起最多数人对于填答问卷的热情,那些在读完你的封面信之后,对你的努力能表示支持的个人、群体或某单位成员对你来说,都是难能可贵的。
- 仔细考虑在 5.5 部分所探讨过的每项技术,以增加问卷的回收率,确定如何才能得到尽可能多的回收问卷,参考上述各问题中所举的例子,关注开放式与封闭式问题的表述方式。表 5.5 是关于提高回收率的各种方式的归纳,其具体操作则要在 5.5 节里探讨。

5.4.3 利用户外资源建构问卷

目前,信息准确成为越来越重要的要求,同时,问卷设计也应坚持节约原则,这使得许多机构开始提供在线使用的问卷,这些问卷一般被安排在因特网上的站点里。

例如,"问卷设计者"(www. questionbuilder. com/)允许人们在网上建构问卷,然后通过某个站点将其发布,或者通过 E-mail 供人分享,像这种技术可以用在基于网络的调查、客户问卷调查、在线研究、数据收集、网站评估、市场分析、选票调查以及 E-mail 调查等方面。

参考文献

Dillman, Don A. (1978). Writing questions. Chap. 3 in *Mail and telephone surveys: The total design method.* New York: Wiley-Interscience.

Lazarsfeld, Paul F. (1972). *Qualitative analysis: Historical and critical essays.* Boston: Allyn & Bacon.

表 5.5 提高问卷回收率的技术

技 术	可能提高的百分比	最好的情况
跟踪(追加访问)	50	使用一种以上的追加访问技术是很必要的,对于至关重要的问题,在追加访问时使用双明信片(double postcards)可以提高回收率,打跟踪电话也很有效。调查者还应该知道填答者是否需要再邮给他们一份问卷(原有问卷可能会丢失,或被弄坏),Sewell 和 Shaw 汇报他们对威斯康星高级学校的学生家长连续使用过三次邮寄和一个电话访问,结果在总共 9 007 份问卷中获得 87.2% 的回收率。(见 1968 年 4 月《美国社会学评论》第 33 期,第 193 页)
(调查)发起人	17	John. K. Norton 发现,如果填答者熟悉调查发起人,那会产生最好的结果,一个由国家机构发起的调查也能收到较好的效果,其余情况如果由高到低排列,依次为:在与调查内容相似领域内工作的地位较低的人、某个出版公司、大学的教授或学生、某个私人组织或基金会。
问卷长度	22	问卷如果较短,那么越短越好,这种情况下,使用双明信片能得到很好的效果。如果问卷内容至少超过了 10 页纸,那么问卷长度就不会对回收率产生太大的影响,这也是 Sewell 和 Shaw 在一次使用双明信片的研究中得出的结论。
介绍信	7	使用一种看来有利于他人的请求方式,比传递调查者将获得一些礼物之类的信息所获得的效果更好。
问题类型	13	那些客观性问题所得回收率最高,而主观性问题所得回收率要差些。

资料来源:检索自网站 www. ncspearson. com,2001 年 8 月 20 日,经允许有改动。

提高问卷回收率的技术* 5.5

NCS Pearson 在网站 www. ncs. com/ncscrop/top/wedo/index. htm 上建议,要提高问卷回收率,可以从 www. ncspearson. com/ncscrop/reserch/97-1. htm 上下载有关信息。

5.5.1 设计好问卷形式

提高问卷回收率最简单易行的途径之一,就是使用那种设计精美,有吸引力,并易于填答的问卷。一般说来,让调查表或问卷更容易填写,是能增加该问卷被回收的可能性的。

* 本部分内容改编自一篇题为《提高回应率》(*Increasing Response Rates*)的论文,本文首次发表于 1997 年 6 月,2001 年 8 月 20 日作了修订。使用获得了授权。

5.5.2 辅助邮寄物

辅助邮寄物包括这样一些事项:事先通知单、委托填表卡、提醒式明信片及重邮调查表(问卷)。

事先通知单与委托填表卡。事先通知单往往是一张明信片或一封信,它对即将邮出的调查问卷做一些解释,并恳请对方参与调查,委托填表卡请求准填答者回邮一张明信片,以指明他们将参与该调查,最近的研究结果(Duhan & Wilson,1990)表明,使用事先通知单可以使回收率达到32.2%,而委托填表卡则可达到20.4%。

提醒式卡片。提醒式卡片一般在初次问卷邮出约一周后发给调查接受者,该卡所起的作用是:对调查做出回应的人表示感谢,对未做出回应的人表示提醒,从使用该卡的经验来看,他们非常有效,可以让作出回应的人由初次邮出问卷之后再提高约10%。

重邮调查表。重邮调查表就是第二次发送问卷调查表,有时是发给所有人,有时则只发给那些没有回应的对象,有两项研究表明,通过这种手段能取得富有成效的结果,在一项研究中(Von Reisen,1979),"重邮"清楚地使得回收率由39.5%提高到50.0%,在另一项研究中(Etzel & Walker,1974)则从38.8%提高为52.9%。

5.5.3 邮件类型

事实证明,不管是发送还是回收问卷调查表,其使用的邮件类型都会对问卷回收率产生影响。对于不同邮件类型以及邮寄方式的研究业已指明,它们对问卷回收率的确带来变化(尽管在这些研究中往往还有其他变量在起作用)。

对于较长的问卷来说,带回执的邮件显得比较有价值。在一项有关回执邮件效力的研究中(Ford,Mclanghlin & Wiuiamson,1992),人们使用带回执邮件来检验不同长度问卷的回收效果,结果表明,带回执邮件这种方式对于较短的问卷几乎没产生什么影响,但对于较长的问卷,使用带回执邮件比使用普通邮件,可以得到近于双倍的回收率(即33.3%比之于17.8%)。人们发现,邮政快件也能产生积极效果,特别是对从事行政管理与商务的填答者更为有效。在一项研究中(Kauis & Gigleriano,1992),当人们在发送和回收问卷中同时使用快件,其回收率达到了52%,相对来说,若使用美国普通信件,并以团体名义寄出或回收时,其数量只能达到26%。尽管邮政快件比美国普通信件在花费上要大得多,但如果问卷调查内容十分敏感,人们预计回收率会很低的时候,它还是适于运用的。可以理解,使用邮政快件做调查,可以增强被调查者对该调查的重要性认识。

最后,资料表明,把回邮信封一并寄给填答者,而不是让他们自己去花钱回邮,也能极大地提高回收率,另外,人们还发现,在信封上贴邮票比在上面盖商务回函印戳效果要好些(不过,这要看邮递的问卷数量,有时候,从代价上看,通过在回邮信封上贴邮票来提高回收率就显得不值)。

Kanuk & Berenson(1975)以及Linsky(1975)认为,封面信是能影响填答者并提高其参与回答可能性地间接方式之一。他们建议应该对封面信给予极大关注,但不巧的是,他们对于如何提高封面信质量没有提供详细的说明。一封封面信应该解释该项调

查的重要性,并指出它对于填答者所具备的潜在价值。尽管 Linsky(1975)报告过一项研究,指出在解释了调查目的并说明了调查重要性之后,其问卷回收率增加了(由29.8%到42.5%),但是,他也报告过在做出相似努力之后,回收率反而降低的情况。他还指出,有时研究者发出需求帮助的恳请只会导致事与愿违。

在 Heberlein 和 Baumgartner 的评论(1978)中(虽然并不是要特别谈论封面信),他们发现,(封面信)有无特色是对调查产生极大影响的因素,特色越明显,回收率就越高。设计封面信,提供有关调查所具备价值的信息,将增强调查的特色,并提高回收率。在 Harvly 的评论(1987)中,总结了封面信的影响,指出其不确定性。特别是他提到过两个案例,在其中调查者恳求帮助并提高了回收率,不过,他也提到过一项研究,调查者同样请求帮助却反而减低了回收率。最后要说的是,Dillman(1991)认为,封面信在提高问卷回收率中有重要地位,可惜他没有提供数据说明,也没有对其操作作出评价。

5.5.4 电话跟踪访问

在访问人员得到较好培训的情况下,电话跟踪访问能够有效地提高回收率。访问人员的电话接触最少能说明调查是重要的,而其沟通也能增强被访者的投入热情。不太好的是,在提高回收率的各种方法中,它的代价显得有点偏高。

5.5.5 奖励物

运用奖励物在问卷回收率的研究中得到极大的重视,虽然人们在事后分析其作用时,往往会得出很不相同的结论。目前出版的有关参考文献中还是就两方面问题形成较为清晰的认识:首先,在邮寄调查法中,运用奖励物是能有效提高回收率的;其次,只是承诺而非马上给予的奖励物,其效果不如随邮给填答者的奖励物。虽然这两种发现已具备很大程度的普遍性,但人们对奖励物能起到的成效范围还是不甚明了。

一些有关奖励物"价值角色"的实际证据存在相互冲突,有些作者提出在预先给予的奖励物中,每增加1美分的投入就可使回收率增长1%,另外一些人则认为,价值提高与回收率增长之间与其说是一条普通上升直线,倒不如说是一条收缩线(diminishing return)。最后,还有一份文献提供这样的数据:在奖励物增值时,回收率既不能显现出直线状态,也没有出现增加。大量数据表明,奖励物增值似乎只能代表某种价值认知。显然,到目前为止,这个问题还远未解决。

社会交换理论是与此讨论相关的具体评价规则体系,该理论认为,人们大多都不愿有欠亏他人的感觉,因而,他们会努力表现出抛却责任、义务等压力的行为。从该理论看,哪怕只是调查时用到的一个极小的标志物,就足以让那些被调查者感觉到某种不得不作出回应的压力,当然,这种标志物不必大得足以让填答者误解为某种雇佣金。

基于这种理论,再参照一些经验研究,我们可以给出一个一般性建议,即0.5美元的奖励物可以使回收率增长到50%,而1美元的奖励物则往往无法使回收率提高,因为这一美元数额已经过多,足以让填答者产生他们已被雇佣,并在领取佣金这种感觉。

参考文献

Armstrong, J. S., & Lusk, E. J. (1987). Return postage in mail surveys: A mete-analysis. *Public Opinion Quarterly*, 51,233-248.

Church, A. H. (1993). Estimating the effect of incentives on mail survey response rates: A meta-analysis. *Public Opinion Quarterly*, 57, 62-79.

Dillman, D. A. (1991). The design and administration of mail surveys. *Annual Review of Sociology*, 17, 225-249.

Dillman, D. A., Sinclair, M. D., & Clark, J. R. (1993). Effects of questionnaire length, respondent friendly design, and a difficult question on response rates for occupant-addressed census mail surveys. *Public Opinion Quarterly*, 57, 289-304.

Dommeyer, C. (1988). How form of the monetary incentive affects mail survey response. *Journal of the Market Research Society*, 30, 379-385.

Duhan, D. F., & Wilson, D. R. (1990, May). Pre-notification and industrial survey response. *Industrial Marketing Management*, 95-105.

Etzel, M. J., & Walker, B. J. (1974). Effects of alternative follow-up procedures on mail survey response rates. *Journal of Applied Psychology*, 59, 219-220.

Ford, R. C., McLaughlin, F., & Williamson, S. (1992, November). Using certified mail in industrial research. *Industrial Marketing Management*, 281-285.

Fox, R., Crask, M. R., & Jonghoon, K. (1988). Mail survey response rate: A meta-analysis of selected techniques for inducing responses. *Public Opinion Quarterly*, 52,467-491.

Harvey, L. (1987). Factors affecting response rates to mailed questionnaires: A comprehensive review of the literature. *Journal of the Market Research Society*, 29,341-353.

Heberlein, T. A., & Baumgartner, R. (1978). Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature. *American Sociological Review*, 43, 447-462.

Hopkins, K. D., & Gullickson, A. R. (1992). Response rates in survey research: A meta-analysis of the effects of monetary gratuities. *Journal of Experimental Education*, 61, 52-62.

Kallis, M. J., & Gigleriano, J. J. (1992, February). Improving mail response rates with express mail. *Industrial Marketing Management*, 1-4.

Kanuk, L., & Berenson, C. (1975). Mail surveys and response rates: A review. *Journal of Marketing Research*, 12,440-453.

Linsky, A. (1975). Stimulating responses to mailed questionnaires: A review. *Public Opinion Quarterly*, 39, 82-101.

London, S. J., & Dommeyer, C. J. (1992, August). Increasing response to industrial mail surveys. *Industrial Marketing Management*, 235-241.

Schlegelmilch, B. B., & Diamantopoulos, A. (1991). Pre-notification and mail survey response rates: A quantitative integration of the literature. *Journal of the Market Research Society*, 33,243-255.

Von Reisen, D. R. (1979). Postcard reminders vs. replacement questionnaires and mail survey response rates from a professional population. *Journal of Business Research*, 7, 1-7.

Yammarino, F. J., Skinner, S. L., & Childers, T. L. (1991). Understanding mail survey response behavior: Meta-analysis. *Public Opinion Quarterly*, 55, 613-639.

Yu, J., & Cooper, H. (1983). A quantitative review of research design effects on response rates to questionnaires. *Journal of Marketing Research*, 20, 36-44.

关于影响问卷回应率因素的推荐读物

Altschuld, James W. , Thomas, Phyllis M. , McColskey, Wendy H. , Smith, Dennis W. , Wiesman, D. , & Lower, M. (1992). Mailed evaluation questionnaires: Replications of a 96 percent return rate procedure. *Evaluation and Program Planning*, 15, 239-246.

Biner, Paul M. , & Kidd, Heath J. (1994). The interactive effects of monetary incentive justification and questionnaire length on mail survey response rates. *Psychology and Marketing*, 11, 483-492.

Boser, Judith A. , & Clark, Sheldon B. (1992, April). *Desirable mail questionnaire characteristics in teacher education research.* Paper presented at the annual meeting of the American Educational Research Association, April, San Francisco.

Christensen, Maribeth. (1997). An interdisciplinary theoretical framework for the mailed questionnaire process and the development of a theory on immediacy and salience as significant variables of response rates. *Dissertation Abstracts International*, 57, 4779B.

Cole, Cornette, Palmer, Randall, & Schwanz, Dennis. (1997). *Improving the mail return rates of SASS surveys: A review of the literature* (Working paper series). Washington, DC: U. S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Crompton, J. L. , & Tian, Cole S. (1999). What response rate can be expected from questionnaire surveys that address park and recreation issues? *Journal of Park and Recreation Administration*, 17, 60-72.

Day, Neil Atherton, Dunt, David R. , & Day, Susan. (1995). Maximizing response to surveys in health program evaluation at minimum cost using multiple methods: Mail, telephone, and visit. *Evaluation Review*, 19, 436-450.

Enger, John M. , Shain, Russell E. , Manning, Tom, Talbert, Lonnie E. , & Wright, Don E. (1992, November). *Response rate effects of three questionnaire formats.* Paper presented at the annual meeting of the Mid-South Education Research Association, Knoxville, TN.

Enger, John M. , & others. (1993, April). *Survey questionnaire format effect on response rate and cost per return.* Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

Fisher, M. R. (1996). Estimating the effect of nonresponse bias on angler surveys. *Transactions of the American Fisheries Society*, 125, 118-126.

Groves, Bruce W. , & Olsson, Roy H. , Jr. (2000). Response rates to surveys with self-addressed stamped envelopes versus a self-addressed label. *Psychological Reports*, 86, 1226-1228.

Hare, S. , Price, J. H. , Flynn, M. G. , & King, K. A. (1998). Increasing return rates of a mail survey to exercise professionals using a modest monetary incentive. *Perceptual and Motor Skills*, 86, 217-218.

Harvey, Lee. (1987). Factors affecting response rates to mailed questionnaires: A comprehensive literature review. *Journal of the Market Research Society*, 29, 341-353.

MacGregor, Elizabeth, & McNamara, John R. (1995). Comparison of return procedures involving mailed versus student-delivered parental consent forms. *Psychological Reports*, 77, 1113-1114.

Newton, Kate, Stein, Samuel M. , & Lucey, Clare. (1998). Influence of mailing strategies on response to questionnaires. *Psychiatric Bulletin*, 22, 692-694.

Paolo, Anthony M., Bonaminio, Giulia A., Gibson, Cheryl, Partridge, Ty, & Kallail, Ken. (2000). Response rate comparisons of e-mail- and mail-distributed student evaluations. *Teaching and Learning in Medicine*, 12, 81-84.

Parthasarathy, Anuradha, & others. (1995, April). *Effects of respondents' socioeconomic status and timing and amount of incentive payment on mailed questionnaire response rates.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Shahar, E., Bisgard, K. M., & Folsom, A. R. (1993). Response to mail surveys: Effect of a request to explain refusal to participate. The ARIC Study Investigators. *Epidemiology*, 4, 480-482.

Vogel, P. A., Skjostad, K., & Eriksen, L. (1992). Influencing return rate by mail of alcoholics' questionnaires at follow-up by varying lottery procedures and questionnaire lengths: Two experimental studies. *European Journal of Psychiatry*, 6, 213-222.

Westcott, S. Wickes, III, & others. (1995, May). *Evaluating reasons for low response from mail surveys: AIR 1995 annual forum paper.* Paper presented at the annual forum of the Association for Institutional Research, Boston.

5.6 实地研究中个人会谈技术的选择与运用

会谈是访谈者与被访者之间的个人接触，一般在被访者家里或办公室进行，从预先设计好一系列问题并进行高度结构化的会谈，到完全是非结构化的，只在访谈者内心存有某些讨论意向的非正式谈话，会谈方式可以说是多种多样。访谈时，其话题的自由程度对访谈结果来说，既充满机遇，也有一定风险。机遇在于人们可以就一些主题深入探究，而风险则在于访谈者可能得不到合适的资料。其实，很多会谈资料常常难于整理与比较。

研究者往往不希望每个开放性问题都占去大量访谈时间，因此，对于每个开放性问题就需要某种指导准则，对每个问题的每个准则，都需要仔细地独立审查，以确定其有效性与可信度。如果研究者不得不用到开放性问题，他就只能小心挑选其中一些，并在内心对这些问题的研究目的保持明确认识。如果要检验某种假设，研究者就要明确：这些问题是应该与假设直接相关的。在需要揭示人们的态度、反应性结构，以及在需要确定某些与社会系统或个性结构内在关联的信息之时，开放性问题是比较适合而且很有效力的。

一般情况下，会谈应保持在 45 分钟的时限之内，“公众舆论会谈”曾经有报告指出，在这一时间点上，大多数被访者开始困倦并对访谈失去兴趣。当然，也有人说在整个访谈过程中一些被访者会被“唤醒”，这也许是真的，在一些文章里就提到 6 小时或 8 小时的会谈例子，不过，这些时间特别长的会谈是特例，只能在特殊准备的环境里才能进行。

访谈有三种形式：结构式访谈（structured interview）、典型（焦点）访谈（focused in-

terview)和自由访谈(free story)。所有这些形式及其特征都在5.1节的表5.1中列出来了。在它们之中,可能会运用到的技术,包括以量表来测量社会因素、态度和个性特征等,其中,匿名投票以及小组活动等方法被经常采用。

下边的内容分别说明个人访谈的优点与缺点,它可以被当作一种核对表来使用,研究者可以给那些重要的、根本的优点添上加号,而对那些消极影响因素添上减号,这样,在是否选择个人访谈这种方法时,人们就有了一个比较可靠的参考基点。

其实,实地研究方法也是可以利用的,它们包括:集体访谈、电话访谈、案例研究、生活史研究、直接观察、参与观察以及大众观察,我们的建议并非为所有这些方法准备的,但在本节末尾,我们还是设计了一个参考表,在表中较为详尽地探讨了所有这些方法与技术。

研究者应针对自身的研究,考察其具有重要价值的优点,其无法克服的缺点,并对最终选择作出评估,他接下来就应该考虑表5.1中所指出的,包括内容分析、邮寄问卷、电话访谈、观察或其他某种方法等内容了。

5.6.1 个人访谈的优点

研究者们已经确认,个人访谈具备一定的优点。我们将其中的14条列举如下:

1)个人访谈通常有很高的回收率,因为大多数人在这种情况下愿意合作。

2)个人访谈可以组建与大众群体非常接近的精确样本(如果被访人员挑选很恰当的话),因为按照这种方式,实际上每个人都可以被找到并进行访谈。

3)这种方法所获取的资料比其他方法更为精确,因为访问人可以向被访问者解释,以澄清那些看来不怎么牢靠的回答内容。甚至,如果被访者故意给出错误回答,访问人也可以立即发觉并利用其他特别手段去弄明真相。

4)访问人可以现场收集一些与被访者个人特性以及周围环境相关的辅助参考资料,这些资料可借助来解释访谈结果,也可用以评估被访者的样本代表性。

5)个人访谈时,可以运用评分或测试等手段,这种近似于实验的方法有利于建立与访问主题相关的精确记录。

6)也可以运用视觉媒体材料以观察被访问者的反应。

7)可以按计划继续回访,询问那些尚未问完的问题,或者核查一些失误的地方,这些一般都不会让被访者产生厌恶之情,从而,与其他方法相比,能确保更大的回访量。

8)被访者对于访问更多做出即刻反应,而若运用邮寄问卷法,他们则会在深思熟虑之后再做回答。

9)在个人访谈时,访问人对于由谁来回答问题是可以控制的,而在邮寄问卷的条件下,家庭主人往往会在填答之前与其他一些人进行商量。而且,如果情况允许,在个人访谈中,也能进行小组讨论。

10)个人访谈时,能给被访者充分的时间,让其接受被访的主题,这样,他们在回忆一些相关经历时也变得很容易。

11)在个人访谈中,访问人可以将一些可能引起被访者敏感的话题有技巧地夹在

谈话中，而且，通过观察被访者的反应，访问人在必要的时候，可以更改话题，或作出进一步解释（如果被访人看起来对该话题有所抵制）。也就是说，与其他方法相比，个人访谈法能对访问情境进行更为有效、更细致的调控。

12）在这种条件下，如果从被访者身上较难得到（也就无从记录）所需信息、资料，我们可以考虑多占用他一点时间。

13）在预定访问安排表无法实施时（参见下面所述缺点 2），可以改换访问时的提问方式，以贴近被访者的能力或教育水平。由此，可以相对减少误解或误导性问题。

14）访谈内容的多少不会影响拒答率。

5.6.2 个人访谈的缺点

与上述优点相比，个人访谈有以下缺点：

1）伴随个人访谈的整个操作过程，会有很大的开支，比如，薪水、交通费用等。

2）已经有访问接受率较低的一些相关报告，特别在大都市里，其居民因为当地的团伙犯罪或贫穷犯罪而改变了生活方式，通往住户的中间过道往往紧紧锁着，一些多单元住房内的居民则因为普遍存有的、不愿接见陌生人的观念而无法访问，甚至在犯罪率较高的生活区，越来越多的人们以沉默来抗议让陌生人进入独居家庭。当然，也有一些访问员不愿到那些被认为有危险的地方去做调查。总之，在大都市，那些住在当地并愿意接受采访的人，其比例最终被认为已经减少到大约 50%（Groves & Kahn，1979）。

有了这种不利状况，人们就更为关注电话访问（下一节将做详细介绍），甚至也有人提出，是否可以把多种收集资料的方法，比如邮寄问卷法、电话访问法、个人访谈法等组合起来加以使用。举例来说，如果已经运用了电话访问法和邮寄问卷法，而仍不能达到所要求的回收率，那么，就可以接着运用个人访谈法，以使样本尽可能齐全，回收率尽可能提高。

3）除非访问员得到正确的培训和控制，否则收集到的资料就会不准确、不完整，一些质量低劣的数据将导致很大比例的访问资料完全没法利用，其比例有可能超过让被访问者自主填答，再回邮给调查负责人这种方式的比例。

4）假如某个对象属于我们要调查的样本中的一员，而我们又可以利用电话沟通，那么个人访谈通常比电话访谈要花费更多的时间，虽然，电话访问在某些方面还不能完全取代个人访谈（比如访谈时，人们彼此面对，会产生丰富而细微的人际互动行为）。

5）如果访问人是在白天进入被访者家中，那他所遇到的大部分被访者将只能是家庭主妇，如果想调查家庭的男主人，那大部分访问就只能安排在晚上或周末，而晚上访问时又只能利用上 1 ~2 个小时，所以，要运用个人访谈法去调查那些上班人员，就得培训大量的访问员同时开展工作。

6）个人偏见将会使问题发生扭曲。比如，常有某种偏见的访问人，可能会无意间问到一些问题，并促发被访者不得不作出让步性回答，这种状况尤其存在于那些有关观念、态度的研究中。为了阻止这种带偏见色彩的问题，大多数调查他人意向问题的人，都会指导其访问员严格依据访问表上印制好的问题提问。

7）研究人员应该意识到，为调查提供基金的那些机构，一般都不太情愿把其资金用在那些过分依赖于个人访谈的研究方案上。考虑到上述不利之处，特别是有可能出现的高投入，低接受访问率的情况，他们对于这种研究方案的申请往往心存戒备。

参考文献

Groves, Robert M., & Kahn, Robert L. (1979). *Surveys by telephone: A national comparison with personal interviews.* New York: Academic Press.

关于面对面访谈的推荐读物

Baker, Reginald P. (1992). New technology in survey research: Computer-assisted personal interviewing (CAPI). *Social Science Computer Review*, 10, 145-157.

Bongers, I. M. B., & van Oers, J. A. M. (1998). Mode effects on self-reported alcohol use and problem drinking: Mail questionnaires and personal interviewing compared. *Journal of Studies on Alcohol*, 59, 280-285.

Buetow, Stephen A., Douglas, Robert M., Harris, Peter, & McCulloch, Colin. (1996). Computer-assisted personal interviews: Development and experience of an approach in Australian general practice. *Social Science Computer Review*, 14, 205-212.

Cacciola, John S., Alterman, Arthur I., Rutherford, Megan J., McKay, James R., & Janssen, Denise May. (1999). Comparability of telephone and in-person structured clinical interview for DSM-III-R (SCID). *Assessment*, 6, 235-242.

Caldwell, David F., & Burger, Jerry M. (1998). Personality characteristics of job applicants and success in screening interviews. *Personnel Psychology*, 51, 119-136.

Collesano, Stephen Paul. (1986). Personal and telephone intervicws: Comparison for trend data. *Disseration Abstracts International*, 47, 329A.

Lynn, Peter, Turner, Rachel, & Smith, Patten. (1998). Assessing the effects of an advance letter for a personal interview survey. *Journal of the Market Research Society*, 40, 265-272.

Reuband, Karl Heinz. (1992). On third persons in the interview situation and their impact on responses. *International Journal of Public Opinion Research*, 4, 269-274.

Schwarz, Shirley P., & others. (1991). Reasons for changing answers: An evaluation using personal interviews. *Journal of Educational Measurement*, 28, 163-171.

Wilson, K., Roe, B., & Wright, L. (1998). Telephone or face-to-face interviews? A decision made on the basis of a pilot study. *International Journal of Nursing Studies*, 35, 314-321.

电话访问前的准备工作概述 5.7

伴随着工作效率的提高，社会科学研究者正越来越多地利用电话进行社会调查，电话访问出现的主要原因在于以下几个方面：

- 在美国几乎所有的居民家中都装有电话。据统计，在横向的抽样调查中，几乎

100%的调查都可以通过电话进行。电话的使用使得调查中社会阶层之间产生偏见的概率大大降低(Dillman, 1978)。

- 当面访问的费用比较昂贵,而电话访问的成本相对较低,在许多情况下,它的花费相当于当面访问的45%左右(Groves,1980)。
- 在大城市,当面访问的回答率逐渐下降,电话访问具有许多优势,它的回答率比当面访问的回答率仅低5% ~10%(Groves,1980)。

电话应用于调查研究的时间较短,这意味着社会科学领域中有关电话访问的知识相当少,以下就是有关电话调查的未知之处:①各种各样的被调查者的回答率;②不同被调查者愿意接受电话访问的时间长短;③与邮寄填答和当面访问相比较,电话访问中的特殊要求。

5.7.1 电话访问指导

电话访问与邮寄问卷法形成鲜明的对照。电话访问完全依赖口头沟通交流,与面对面的访问调查不同,访问者和被访问者必须在互不见面的条件下进行互动。电话访问者不能借助视觉解释问题,也不能观察到被访问者表情中暗示的误解。

电话访问问卷的设计必须考虑到满足三类群体的需要:被访问者、访问者、编码者。

(1)被访问者

让人接受电话访问是有一定难度的,被访者可能偶然间接到电话并被要求做某些事情——回答访问者的综合问题——这些问题他们可能不能完全理解。他们可能正在做其他的事情,例如准备晚餐、玩游戏、读报、听广播,或看电视。他们的情绪各不相同,从挫折、怀疑、焦虑到完全漠不关心都存在。需要一些微妙的技巧使被访者从他们开始或正在做的事情中摆脱出来。因为这些事情会转移他们对访问的注意力。进入电话访问状态意味着被访者需要时间去习惯访问者的说话方式。去理解不同的词汇和问题的含义以及其他的东西。对许多人来说,接受电话访问是一种新的经历。被访者可能需要时间去思考问题的答案,绝大多数的被访问者在被调查过程中需要得到支持和鼓励。

(2)访问员

成功的访问员必须保证电话问卷的完整性。问卷上的信息应该准确无误并且能够准确地记录答案。访问员必须决定被访问者家庭中谁负责回答问题,并使他(她)接听电话,电话访问能否成功进行,开头的几分钟起着至关重要的作用。访问过程中可能会遇到诸如此类问题,如:"我怎么知道你就是你说的那样呢?","你到底在兜售什么?"或者"我对你或者你的产品一无所知,你为什么不访问别人呢?"为了不接受访问员的电话访问或拒绝访问员,他们可能找出一些借口,比如:"我正在等一个电话——你现在不得不挂掉电话,呆会儿再打过来好吗?"

访问员必须证明他们的正当性和课题项目的价值,他们必须鼓励被访者开始接受访问,在访问调查过程中,访问员必须控制访谈过程,逐个问题进行访问。在记下被调查者回答的问题的同时准备问下一个问题。需认真听取并捕捉任何情绪的变化,记录

被访者主动提供的评论话语,握紧听筒转向问卷的下一页。

(3)编码者

掌握资料收集、整理的简便方法是重要的。预先编码可以鉴别问卷上的每种可能回答。事先为编码者做好标记和附加的指示通常不会干扰访问员的访问,其中一些重要的步骤如下:

- 设计电话访问问卷时要尽量使所有问题直接、明了,并按一定顺序认真地排列,要把易回答的、有趣的、社会性重要的,以及与调查研究主题密切相关的中心问题放在问卷开头。要把所有与主题相关的问题放在有关的个人特征问题之前。
- 一份精心设计组织的问卷将最大限度地保证访问员把问卷问题的原样确切地传达到每一个被访者。每一个回答目录被指派一个可鉴别的数字,这个数字在电脑数据资料中用来指代被访问者的回答选择。最终的问卷要做试测,很可能还要进行访问员角色扮演,以及"访问员"和"被访者"角色的互换。
- 对调查访问的实施来说,控制电话访问中最难的环节是如何组合研究设计中的所有部分。每一个准备工作都是为开始调查访问这一关键环节服务的。

以下的活动步骤(Dillman,1978)都是在访问前完成的。这些准备工作将有助于避免一些因疏忽和组织设计不同而造成的失误。

(4)电话访问检查单

以下就是一份简洁的检查单,它有助于研究者进行电话访问前的准备工作。

- 步骤1:选择样本
  - 从电话簿里获得姓名、街道和电话号码,打印成标签附到问卷的封面上。
  - 通过电脑(或通过随机数表)随机选择一些号码,打印出来分发到访问员手里。
- 步骤2:设备和器材
  - 有使用电话的便利条件。在某些情况下,利用蜂窝式电话相对比较便宜,这在一定程度上依赖于长途电话开通的数量。
  - 检查电话确保它们能正常工作。
  - 可以较便利地租用电话线(如有必要)。
  - 放好桌椅(如果必要)。
  - 用纸盒把问卷分成适当的类别并贴上标签(例如拒访、完成和复访)。
- 步骤3:与电脑相关的必需工具(如果需安排即时数据资料的处理)
  - 筹备好一些电脑(包括便携式电脑和个人数字式辅助器,如掌上电脑等)。
  - 决定将要利用的分析程序软件和使用说明书。
  - 运行电脑,利用分析程序对"仿真"数据进行查错功能调适。
- 步骤4:材料
  - 问卷设计。
  - 问卷复印。
  - 问卷装订。
  - 问卷封面上附加的选择程序(如有必要)。

  · 附在封面上的一份姓名地址目录单(如果可得到的话)。
  · 问卷随机分发到访问员。
  · 复印访问规则说明书。
  · 复印特别的拨号指令说明(如有必要)。
  · 铅笔、笔记本、图钉、橡胶带以及其他各种辅助工具。
  · 在每一个访问点都要附有以上的材料工具。
- 步骤5:预先通知信或联系信
  · 把通知信打印出来,填写好有关内容后,放进信封里。
  · 电话访问前3~5天邮寄出信件。
- 步骤6:人员组成和培训安排
  访问人员和编码人员是:
  · 雇来的。
  · 经过培训的。
  · 预先安排好的。
  · 事先安排好监督人员。
  · 预先安排人员对问卷填答的完整性进行核查。
- 步骤7:其他资源和必需工具
  · 获得调查地区的电话簿(以便帮助查错、纠错)。
  · 把调查进度安排通知有关人员。

参考文献

Dillman, Don A. (1978). Mail and telephone surveys: The total design method. New York: Wiley-Interscience.

Groves, Robert M. (1980). Telephone helps solve survey problems. *Newsletter of the Institute of Social Research* (University of Michigan), 6(1).

关于电话访谈的推荐读物

Bachofen, Martin, Nakagawa, Akiko, Marks, Isaac M., Park, Je Min, Greist, John H., Baer, Lee, Wenzel, Keith W., Parkin, Richard J., & Dottl, Susan L. (1999). Home self assessment and self-treatment on obsessive-compulsive disorder using a manual and a computer-conducted telephone interview: Replication of a U. K.-U. S. study. *Journal of Clinical Psychiatry*, 60, 545-549.

Blair, Ed. (1989). An experimental comparison of telephone and personal health interview surveys. *Public Opinion Quarterly*, 53, 145-146.

Bowen, Gary L. (1994). Estimating the reduction in nonresponse bias from using a mail survey as a backup for nonrespondents to a telephone interview survey. *Research on Social Work Practice*, 4, 115-128.

Fournier, Louise, & Kovess, Vivianne. (1993). A comparison of mail and telephone interview strategies for mental health surveys. *Canadian Journal of Psychiatry*, 38, 525-533.

Greenfield, Thomas K., Midanik, Lorraine T., & Rogers, John D. (2000). Effects of telephone versus face-to-face interview modes on reports of alcohol consumption. *Addiction*, 95, 277-284.

Katz, Elinor. (1993, April). *A critical analysis of interview, telephone, and mail survey designs*. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

Maynard, Douglas W., & Schaeffer, Nora Cate. (1997). Keeping the gate: Declinations of the request to participate in a telephone survey interview. *Sociological Methods and Research*, 26, 34-79.

Moore, Robert John. (2000). Achieving understanding in the standardized interview: A conversation analytic study of talk in telephone surveys. *Dissertation Abstracts International*, 60, 3142A.

Tran, Thanh V. (1995). Telephone interview timing and measurement properties of well-being in a sample of elderly Hispanics. *Journal of Social Service Research*, 20, 29-47.

Wester, Fred. (1996). Telephone polls in social research: Methods, techniques, interview practice. *Communications*, 21,507-508.

电话调查与当面访谈的比较 5.8

电话访问和通过随机拨号抽取家庭样本程序的发展也许是自多阶段概率抽样以来调查研究方面最重要的革新了。电话调查与当面访谈的比较提出了一些重要的问题。

(1)电话调查与当面调查的有效性比较

即使在所有抽样条件和调查内容稳定不变的情况下——而这几乎是不可能的，这个问题也没有简单的答案。但是，在诸如抽样覆盖率、家庭中被访者的选择、总回答率和拒访率、回答的效度等方面，确凿的数据资料都是有用的。

几乎100%的美国居民都可以通过电话进行接触，因此电话抽样调查的覆盖率开始接近概率抽样中当面访谈的覆盖率水平。

入户抽样调查过程中对被调查者的选择必须在访问的开始阶段完成。在当面访谈过程中，对访问者的选择主要利用家庭成员名单完成，在电话调查中，这个选择过程可以简单地通过一个与不同数目的成年男性与成年女性相对应的栅格来进行。电话访问员先确定被访家庭中成人的总数和妇女的数目，然后使用栅格指出被选作回答者的那个人。这一选择过程的使用表明，它在样本户中对被调查者的选择会出现大约10%的误差(Groves & Kahn,1979)。

全国性的电话调查的回答率至少比当面访谈的回答率低5%。不论访问员的培训、控制和反馈程序，以及向被访者的介绍技巧等许多方面有多大变化，它们之间的这种对比关系是相当稳定的(Grove, Lyberg,& Massey,1988)。

(2)电话调查与当面访谈中，被调查者提供答案的内容或深度的系统差异比较

一些研究者指出，电话访问和当面访谈这两种调查方式在完成调查的速度上存在类似的差别。电话访问的速度与被访者对问卷结尾开放式问题的回答数量和类型的差异有关。一些人接受电话访问但并不一定能完成整个访问过程。电话访问被认为回答率较低(低5%)。尽管存在这些差别，但这两种资料收集方式在回答上的差异极少有统计上意义。

有些研究集中在有关敏感资料或令人困窘的问题的报告上，这些研究大多发现在电话访问和当面访谈之间不存在或仅存在非常细微的差别。许多研究结论不相一致是由于在研究设计、所调查的人群或资料收集方式上存在差异。总体的陈述也是不合适的。

真正受到关注的是电话调查中访问员和被访者之间和谐融洽的关系。无回答表明被访者把电话访问看成是更少报酬的经历和比当面访谈更烦人的事情。在以后的研究中应优先注意研究电话访问的技巧，以激发被访者的兴趣，建立与当面调查同等的信任程度。

(3)电话调查与当面访谈的效率比较

在全国性的社会调查中，电话调查所需人员小于同等规模的当面访谈的人员数量。打个比方，当面访谈调查需要200访问员，每个访问员访问7~8户家庭，在同样的时间里，电话调查只需30~40个访问员，每个人可访问40~50户。监督管理人员和协调人员相应减少。很明显，电话访问存在成本优势。

电话访问的高效率来自于访问后对问卷的立即编码，编码的过程中访问员所犯的错误可很快被发现并被纠正，访问员的访问行为能迅速获得反馈，访问结果也能及时的得到。

抽样越来越多地采用随机数字拨号的方法，这种方法是在电话簿上直接选择号码。一般来说，一组随机选择的数字对应于一组工作电话号码。这意味着随机抽样方法更简单更快速，甚至可以起到事先通知的作用。

电话用户接听电话时，首先听到的是对调查的介绍录音，然后是调查员为收集真实确切资料的提问声音。

表5.6 对邮寄问卷法、当面访谈和电话调查三者的比较

影响覆盖率和信息资料可靠的因素	邮寄问卷	当面访谈	电话调查
最低的相对成本	1	3	2
最高的回收率	3	1	2
最高的资料可信度	2	1	3
最大样本覆盖范围	3	1	3
完整性，包括敏感资料	3	1	2
总体的信度和效度	2	1	3
获得资料要求的时间	3	2	1

注：1=最令人肯定的等级　2=中间位置的等级　3=最少令人肯定的等级

参考文献

Groves, Robert M., & Kahn, Robert L. (1979). *Surveys by telephone*. New York: Academic Press.

Groves, Robert M., Lyberg, Lys, & Massey, James. (1988). *Telephone survey methodology*. New York: John Wiley & Sons.

在邮寄问卷法、当面访谈和电话调查三者之间作出选择 5.9

对于收集资料方式的选择涉及很多因素。对结果的妄加猜测是不恰当的,因为从某种意义上说,每次的抽样以及其周围的环境条件都是独一无二的。即使对所有事件进行严格的控制,变化仍将不可避免。如果研究者准备进行许多的后续工作或结合多种方法进行资料收集,那么就可预期有更高的回收率。

表5.6展示了对不同资料收集方式在7个重要因素上的评估。在选择研究设计时,研究者必须确定在这些因素中哪些是最重要的,然后选择其中一种方式,同时把它的优点和缺陷记在脑海里。

当然成本总是一个优先考虑的因素。表5.6显示邮寄问卷花费最少,电话调查花费次之,当面访谈成本最高。但是如果电话调查包含大量的长途电话,相比较当面访谈而言,花费成本因支出电话费而增多,其成本优势大减。

比较邮寄问卷和当面访谈,会发现,二者的成本花费和其他几乎所有的因素成反比关系。与其他的方式相比较,如果不考虑其最薄弱的环节,当面访谈在很多方面都处于领先地位。电话调查在一系列评估因素的排行上都处于中间位置,这就是电话调查不断得到普及的原因。

无论什么时候,只要条件可能,当面访谈是研究者的首选方式,这是合情合理的。但是邮寄问卷和电话调查仍有其自身的优势。如果要减少调查的成本、预约的被访者比较忙而难以接触到、问卷的问题比较敏感或者涉及个人的隐私、调查时间比较长或内容比较复杂时,邮寄问卷方式是更可取的;当需要很快获得调查结果,访问者需通过训练,被调查的样本较小时,电话调查的方式更为可取。

关于调查的推荐读物

Faria, A. J., Dickinson, John R., & Filipic, Timothy V. (1990). The effect of telephone versus letter prenotification on mail survey response rate, speed, quality and cost. *Journal of the Market Research Society*, 32, 551-568.

Frey, James H. (1983). *Survey research by telephone*. Newbury Park, CA: Sage.

Leftwich, Wade. (1993). How researchers can win friends and influence politicians. *American Demographics*, 15, 9.

Mohebi, Ed, & Hechter, Edward. (1993). Telemarketing: Faster, better, cheaper. *Bank Marketing*, 25, 36-38.

Okun, Jill. (1987). Lead generation media usage in business-to-business. *Direct Marketing*, 50, 142-146, 208-211.

Riche, Martha F. (1990). A bigger role for telephone interviews. *American Demographics*, 12, 17.

Savini, Gloria. (1989). When two technologies meet. *Direct Marketing*, 51, 54-58.

5.10 作为研究工具的同组调查方法

社会科学家目前对涉及时间变化的数据资料的分析方法给予了极大的关注。这意味着所有的社会科学方式已经从静态模式转向动态模式。在动态模式中,同组研究方法是一种重要的研究工具。

同组研究方法(panel technique)指在两次或多次不同时段场合对同一群人进行调查访问。它主要通过重复地访问调查一批人,研究人们行为或态度的变化过程。大多数的同组研究都包含一套核心问题或观察,这些核心问题或观察几乎在每一次调查中都被使用。在每一次或不同次调查中可能获得相当多的补充材料,这可用于解释被访者在核心问题上发生的变化。

同组研究可用于许多领域的调查,如政治民意测验、职业与收入发展趋势的变化过程、消费习惯、大众传播媒介等。同组研究方法主要有两种:一种是集中研究人们的态度和观点,另一种则主要是处理一些经济、消费、沟通交流和其他行为的事实材料。

集中研究人们态度和观点变化的同组研究总体上只需数量很有限的一组人,通常是2~4人,很少超过7人,因为他们想通过被揭示出的变化过程详细确切地说明某一特殊事实或某一原因,他们通常局限于研究人们具体而又明确的社会态度在短期范围内的变化。

5.10.1 同组研究方法的优点

研究者在探究人们某一特定态度在短期范围内的变化有两种选择。首先,通过与被调查者的一次接触交流最终可以获得有关事实或事件经历的资料,可能问到诸如"一个月前你曾经打算投票支持谁?"和"最近你换工作了吗?"等问题。这些获得的信息可能是不准确的,因为被访者可能不能正确地回忆,访问员所问的问题可能易导致误解,或两者兼而有之。建立在重复调查访问基础上的同组研究方法被认为更可靠和更有效,并有许多明显的优点。Zeisel(1957)已经列举出了这些优点,我们在表5.7中可以看到。

5.10.2 同组调查方法本身存在的困难

采用同组调查方法遇到两种基本的内在困难就是小组样本的大量缺损(panel mortality)和重新访问带来的偏差(reinterviewing bias)。

- 小组样本的大量缺损是指,由于两次或更多次访问时难以找到原来的同组成员,或由于被访者拒绝再次合作而导致的同组成员的缺损。因为小组中不同类型成员可能有不同的缺损率,从而导致出现抽样偏差的危险。

- 重复访问的偏差是指,就某一个主题对被访者的行为或他们的态度进行重复讨论带来的被访者的观点、看法产生变化。

表 5.7 重复性横向调查法与同组调查法的比较

横向调查法	同组调查法
1. 记录、反映变化情况 例如,比较两个不同时期的××牌子肥皂使用者比例的变化,包括每次访问时用户之总体差异:净变化。	除净变化外,可获得有关个人用户在数量、方向、时间上转移的准确图表,这些能解释发生净变化的原因。
2. 变化的原因 比如,是否因某种宣传而改变了一个人的政治态度。这是很难问也很难回答的。	分别分析接触某种宣传的一组人和未接触到某种宣传的另一组人,同组调查可证实这两组人的态度变化的数量和方向是否不同。
3. 信息收集的数量(总值) 因为每次调查的受访者不同,研究者无法了解每个被访者的更多信息。	重复访问同一组被调查者可获得一些持续性的大量信息资料。三次访问的同组研究能够提供两三个小时的造访时间,从而获得有价值的信息。
4. 涉及时间段的数据资料 如果调查问题涉及的时间即是访问进行的时间,那么一次访问调查将获得准确的调查结果。如果被访者需要对经过了一段时期的事件进行回忆,那么他们必须依靠记忆力。如果某一研究需要对某一概念做界定,那么总是需要诸如此类的资料的,比如:如果我们要研究一个人的阅读习惯或他(她)是否是某一电台节目的忠实听众,我们必须完全依赖于被访者的记忆和判断。	如果目的是为了收集有关持续了一段时间的资料,通过同组研究方法就可以避免完全依赖被访者的记忆力。重复访问可以获得有关习惯和态度的波动性、一惯性方面的客观资料,访问员也能比较准确、可靠地区分"正常"与"非正常"之间的差别。
5. 结果的可靠性 不同调查访问中观察资料变化的统计意义取决于某一特定样本的结构和规模。	在相同样本规模和结构的条件下,同组调查观察到的变化比重复进行的跨组调查观察到的变化具有更显著的统计学意义。

资料来源:Hars Zeisel(1957),Say it with figures(4th ed)(New York:Harper & Row) pp. 217-218

(1)小样本多次访问中的成员缺损

随着时间的流逝,必然会导致小组成员的缺损,时间越长,缺损的人数就越多,重复访问失败的主要原因在于被访者暂时不在或住址完全发生变化。只要在第一次访问时记下被访者的家庭地址和工作地点,再次访问时尽力找到原有的被访者,以上的这些问题和困难就可以得到部分的解决和克服。电话通知被访者后,还可以尝试邮寄问卷或当面访谈,附有书面证明信可以帮助提高回答率。

缺损造成的偏差随着同组成员缺损数量的增长而增大,缺损并不是随机发生的。例如,众所周知,和低收入阶层一样,年轻人和大城市居民比年龄大的、居住在小城镇

的和上层社会的被调查者有更高的流动性,研究者必须发现每个个案是否存在这种偏差。这种工作中的大量努力和精力都是花在非被调查者身上。积极有效的努力就是使用各种途径和方法,坚持不懈地保证尽可能多的再访机会。当第二次访问的对象不是第一次访问的对象时,研究者必须仔细分析资料,检查是否产生因成员缺损导致了偏差,以及它对于解释这些结果产生的后果有多严重。

刊登在1986年的《应答分析》(*Response Analysis*)中一篇有关消费者同组研究(Sampler)表明,同组调查访问法在第二次抽样调查时至少保证67%的回收率才有效。这项研究由美国海军为其军人及其家属开办的全球性零售连锁店里的消费者组成,在1986年的头八个月里,研究涉及了两个不同地点,分四个调查阶段,此项研究提出了两个问题:一是,作为同组研究样本的这些消费者参加这项调查研究的意愿有多强?二是,完成三次调查访问的小组成员的百分比有多大?

受过特别训练的调查员分别到两个不同的连锁店寻找愿意参与研究的消费者,94%的被访者表示愿意成为同组成员。在第一轮调查中,总共发出了2 190份邮寄问卷,被访者完成了1 989份,回收率为91%。

第二轮调查显示,一旦小组成员完成了第一份调查问卷,他们更有可能继续这项研究。第二轮调查的回收率达78%。在第三轮调查中,回收率达89%。访问员向所有完成三次调查的每个消费者提供了价值10美元的礼物和证书。组织设计和指导这项研究的副主编阿尔·沃格在分析回收率时认为,在研究中,每个阶段回收率的变化显然是由于受到调查中奖励变化的影响。

5.10.3 重复访问的偏差

偏差可能是由于单纯的重复访问本身造成的,因为第一次调查有可能提高被调查者的注意力,他(或她)也可能花更多的时间对调查主题进行考虑并因此改变自己观点、态度。由于重复访问导致的熟悉程度提高也可能增加被访者看法的偏差。

为了测试这些偏差,研究者可以使用与同组调查小组成员相类似的控制组。在同组调查小组成员第一次接受调查时,与小组成员相同结构的控制组样本也就同一个主题接受调查;同组调查小组成员第二次接受调查时,一个不同的样本作为控制组也接受调查。用这种方式,通过比较小组成员在两次调查中随控制组样本变化而产生变化的情况,能够发现、了解任何结果的偏差。

在同组研究中,两个最主要的误差来源:成员缺失和重复访问偏差需要引起注意。但是,许多研究表明,这两个导致误差的因素中的任何一个都不足以严重威胁到同组方法的使用,在社会研究领域中,同组研究是一种前景十分看好、生命力强大的研究工具。

在设计组织一项同组研究时,需要做出许多重要的决定,在表5.8中将它们简要描述出来了。

表 5.8 同组研究设计中存在的问题及其处理建议

设计中的问题	建 议
1. 调查次数 调查次数必须考虑到被访者的合作程度,由于重复访问,被调查者可能变得厌烦、易怒或急躁。重复访问花费较大,廉价的劳力和省钱的访问是不存在的。	调查次数通常应尽可能少到最低限度。
2. 不同调查之间的时间间隔 调查次数之间间隔的时间长短取决于记忆的清晰度、所期望获得信息的类型以及环境变化的速度等多种因素,如果间隔时间太长,小组成员可能丧失对以往经历事件的记忆,这会导致对问题回答的偏差。	所期望获得信息的类型是决定恰当的间隔时间的最重要因素。 消费同组调查通常的间隔时间在 1 ~2 个周。政治同组调查通常要间隔 1 ~2 个月,在激烈的政治运动或政治危机中,间隔时间应该短些。
3. 样本规模 使用大规模样本可以降低抽样误差,但大样本的使用导致成本的提高。第一次调查前,使用较大的样本可以弥补缺损率带来的偏差(例如,在原来的样本设计中包含 1 200 个人,以便能有 1 000 个人接受调查访问)。这里的问题在于原来那些未被访问者可能不同于被访者群体。	通过不同规模的样本检查抽样误差。 在成本固定不变的情况下,研究者要么重复访问以降低初次访问的样本缺失,要么获得一个初次访问时未被访问者的样本。
4. 样本类型:配额与地区 配额样本(quota samples)比地区样本(area samples)具有更简单、更省钱的优势。如果根据特定配额的要求,在大街上选择被调查者,其调查速度和成本都将会很不错。获得被调查者的姓名和地址后,研究者可以在被访者的家中进行下一次访问。但这并不是很可行,因为由于许多原因导致许多初始样本的缺失。地区样本相对来说更少受到样本缺失的影响。	任何时候,只要可能的话,私人性个案访谈小组应控制在以家庭为基础的区域进行。
5. 奖励手段 为获得被访者的配合和参与,有必要使用一些奖励手段。特别是在同组研究中,如果被调查者长时间配合和多次重复接受访问,这一点更重要。研究者要面临两个重要问题:首先是访问的安全性,接下来是被访者持续的合作。	两种主要的奖励手段,即金钱和精神名望(倡议者、发起者的名义),是否有效,通常取决于研究者(主办者)和同组被访者的类型。 在消费者同组研究中,经常使用物质奖励。 政府倡议或有关教育的同组研究,常常强调声望和作为一个市民的责任,鼓励人们参与推动知识和社会进步的活动。

续表

设计中的问题	建　议
6. 访问类型:当面的、电话的、邮寄的 当面访谈是同组研究中使用最普遍的一种方式,但这也是最昂贵的一种调查方法,而且成本越来越高。电话访问因此变的越来越普遍,因为在家中或工作地点,电话的使用几乎普及,但通过电话访问很难获得许多敏感的信息和有关人们态度的资料。	由于当面访谈成本的提高,电话访问正越来越受欢迎,一些新技术和访问技巧的使用使得电话访问更加可靠和有效。 只要电话访问或当面访谈可行的话,一般不采用邮寄访问。
邮寄访问因调查者与被调查者不能直接对话,故而存在两个问题:一是当同组调查访问间隔时间很重要时,邮寄访问就不太现实,因为邮寄调查无法控制时间因素;二是因为邮寄调查是完全开放式的,无法避免由于被访者家人或邻居参与讨论回答而带来的偏差。	邮寄访问也有一些优点,它能进入城市那些不安全地区进行调查访问,或者获得一些敏感的资料。

参考文献

"Sampler." (1987, September). *Response Analysis*, 47, 2.

Zeisel, Hans. (1957). *Say it with figures* (4th ed.). New York: Harper & Row.

关于同组研究方法的推荐读物

Alwin, Duane F. (1992). Information transmission in the survey interview: Number of response categories and the reliability of attitude measurement. In Peter V. Marsden (Ed.), *Sociological Methodology*. Washington, DC: American Sociological Association.

Alwin, Duane F., & Krosnick, Jon A. (1991). Aging, cohorts, and the stability of sociopolitical orientations over the life span. *American Journal of Sociology*, 97, 169-195.

Arzeimer, Kal, & Klein, Markus. (1999). The effect of material incentives on return rate, panel attrition and sample composition of a mail panel survey. *International Journal of Public Opinion Research*, 11,368-377.

Bamberg, Sebastian, & Schmidt, Peter. (1998). Changing travel-mode choice as rational choice: Results from a longitudinal intervention study. *Rationality and Society*, 10, 223-252.

Barreiros, Lidia. (1995). The European Community Household Panel (ECHP): Its design, scientific and policy purposes. *Innovation*, 8, 41-52.

Choi, Namkee G. (1994a). Changes in labor force activities and income of the elderly before and after retirement: A longitudinal analysis. *Journal of Sociology and Social Welfare*, 21, 5-26.

Choi, Namkee G. (1994b). Racial differences in timing and factors associated with retirement. *Journal of Sociology and Social Welfare*, 21, 31-52.

Choi, Namkee G. (1999a). Living arrangements and household compositions of elderly couples and singles: A comparison of Hispanics and blacks. *Journal of Gerontological Social Work*, 31,41-61.

Choi, Namkee G. (1999b). Racial differences in the contribution of wife's earnings to family income distribution. *Journal of Poverty*, 3, 33-51.

Dex, Shirley, Robson, Paul, & Wilkinson, Frank. (1999). The characteristics of the low paid: A cross-national comparison. *Work, Employment and Society*, 13, 503-524.

Dey, Eric L. (1997). Undergraduate political attitudes: Peer influence in changing social contexts. *Journal of Higher Education*, 68, 398-413.

Dillbeck, Michael C., Banus, Carole Bandy, Polanzi, Craig, & Landrith, Garland S., III. (1988). Test of a field model of consciousness and social change: The transcendental meditation and TM-Sidhi program and decreased urban crime. *Journal of Mind and Behavior*, 9, 457-486.

Dirven, Jan Henk, & Berghman, Jos. (1995). The evolution of income poverty in the Netherlands: Results from the Dutch Socio-Economic Panel Survey. *Innovation*, 8, 75-94.

Ellis, Robert A., & Herrman, Margaret S. (1985). The stigmatization hypothesis re-examined: The social reaction to career-oriented college women. *Sociology and Social Research*, 69, 527-547.

Habich, Roland, & Speder, Zsolt. (1999). Income dynamics in three societies. *Szociologiai Szemel*, 3, 3-27.

Hamilton, Jack. (1994). What is marketing research? *Sotsiologicheski Issledovaniya*, 1, 109-120.

Haveman, Robert, Wolfe, Barbara, & Wilson, Kathryn. (1997). Childhood poverty and adolescent schooling and fertility outcomes: Reduced-form and structural estimates. In Greg J. Duncan & Jeanne Brooks-Gunn (Eds.), *Consequences of growing up poor* (pp. 419-460). New York: Russell Sage Foundation.

Havens, Betty, Hall, Madelyn, & Shapiro, Evelyn. (1998, July). *Self-care, informal care, and formal care in the face of aging and health reform.* Paper presented at the International Sociological Association meetings, Montreal.

Hunt, Marjorie E. (1997). A comparison of family of origin factors between children of alcoholics and children of non-alcoholics in a longitudinal panel. *American Journal of Drug and Alcohol Abuse*, 23, 597-613.

Ivie, Rachel L., Gimbel, Cynthia, & Elder, Glen H., Jr. (1991). Military experience and attitudes in later life in contextual influences across forty years. *Journal of Political and Military Sociology*, 19, 101-117.

Jagodzinski, Wolfgang, Kuhnel, Steffen M., & Schmidt, Peter. (1987). Is there a "Socratic effect" in nonexperimental panel studies? Consistency of an attitude toward guestworkers. *Sociological Methods and Research*, 15, 259-302.

Johnson, David R. (1988). Panel analysis in family studies. *Journal of Marriage and the Family*, 50, 949-955.

Jowell, Roger, Hedges, Barry, Lynn, Peter, Farrant, Graham, & Heath, Anthony. (1993). The 1992 British election: The failure of the polls. *Public Opinion Quarterly*, 57, 238-263.

Kaase, Max. (1986). The micro-/macro-puzzle of empirical social research. Some notes on the problem of aggregate stability with individual instability in panel studies. *Kolner Zeitschrift fur Soziologie und Sozialpsychologie*, 38, 209-222.

Kalton, Graham, & Citro, Constance F. (1995). Panel surveys: Adding the fourth dimension. *Innovation*, 8, 25-39.

Kitson, Gay C., & Morgan, Leslie A. (1990). The multiple consequences of divorce: A decade in review. *Journal of Marriage and the Family*, 52, 913-924.

Klein, Thomas. (1987). A new index for measuring individual versus structural change in panel surveys. *Zeitschrift fur Soziologie*, 16, 145-154.

Krohn, Marvin D., & Thornberry, Terence P. (1999). Retention of minority populations in panel studies of drug use. *Drugs and Society*, 14, 185-207.

Krosnick, Jon A. (1991). The stability of political preferences: Comparisons of symbolic and non-symbolic attitudes. *American Journal of Political Science*, 35, 547-576.

Link, Bruce G., & Shrout, Patrick E. (1992). Spurious associations in longitudinal research. *Research in Community and Mental Health*, 7, 301-321.

Macintyre, Sally. (1992). The effects of family position and status on health. *Social Science and Medicine*, 35, 453-464.

Marsh, Herbert W. (1993). Stability of individual differences in multiwave panel studies: Comparison of Simplex models and one-factor models. *Journal of Educational Measurement*, 30, 157-183.

Mayer, Karl Ulrich. (2000). Promises fulfilled? A review of 20 years of life course research. *Archives Europeenes de Sociologie*, 41, 259-282.

Mayer, Lawrence S. (1986). Statistical inferences for cross-lagged panel models without the assumption of normal errors. *Social Science Research*, 15, 28-42.

Mayer, Lawrence S., & Carroll, Steven S. (1987). Testing for lagged, cotemporal, and total dependence in cross-lagged panel analysis. *Sociological Methods and Research*, 16, 187-217.

Mayer, Lawrence S., & Carroll, Steven S. (1988). Measures of dependence for cross-lagged panel models. *Sociological Methods and Research*, 17, 93-120.

McBroom, William H. (1988). Sample attrition in panel studies: A research note. *International Review of Modern Sociology*, 18, 231-245.

Mecsko, Lisa A., & Dunkelberger, John E. (1995). Influence of aspirations and background characteristics on educational attainments of college agricultural graduates. *A Review of County Health Councils*, 33(6), 1-6.

Mueller, Charles W., & Parcel, Toby L. (1986). Ascription, dimensions of authority, and earnings: The case of supervisors. *Research in Social Stratification and Mobility*, 5, 199-222.

Ribisl, Kurt M., Walton, Maureen A., Mowbray, Carol T., Luke, Douglas A., Davidson, William S., II, & Bootsmiller, Bonnie J. (1996). Minimizing participant attrition in panel studies through the use of effective retention and tracking strategies: Review and recommendations. *Evaluation and Program Planning*, 19, 1-25.

Rose, David. (1995). Household panel studies: An overview. *Innovation*, 8, 7-24.

Schupp, Jurgen, & Wagner, Gert G. (1995). The German socio-economic panel: A database for longitudinal international comparisons. *Innovation*, 8, 95-108.

Scott, Jacqueline. (1995). Using household panels to study micro-social change. *Innovation*, 8, 61-73.

Snyder, Leslie B. (1991). Modeling dynamic communication processes with event history analysis. *Communication Research*, 18, 464-486.

Taris, Toon W. (1996). Modeling nonresponse in multiwave panel studies using discrete-time Markov models. *Quality and Quantity*, 30, 189-203.

Taris, Toon W. (1997). On selectivity of nonresponse in discrete-time multi-wave panel studies. *Quality and Quantity*, 31, 79-94.

Taylor, Marcia Freed, & Schaber, Gaston. (1995). An integrated longitudinal database for comparative analysis: The panel. *Innovation*, 8, 53-59.

Toth, Istvan Gyorgy. (1995). The first two waves of the Hungarian Household Panel: Methods and results. *Innovation*, 8, 109-121.

Trivellato, Ugo. (1999). Issues in the design and analysis of panel studies: A cursory review. *Quality and Quantity*, 33, 339-352.

Wang, Jichuan, & Fisher, James H. (1994). Comments on "Estimating macro-relationships using micro-data: One-stage approach." *Sociological Methods and Research*, 22, 520-531.

Willits, Fern K., & Crider, Donald M. (1988). Transition to adulthood and attitudes toward traditional morality: A two-panel study. *Youth and Society*, 20, 88-105.

Zapf, Wolfgang. (1994). Transformation in the former GDR and the sociological theory of modernization. *Berliner Journal fur Sociologie*, 4, 295-305.

Zeh, Jurgen. (1989). Social control by public opinion. A sociological reflection on an issue in commu-nication science. *Publizistik*, 34, 29-45.

社会量表和指标的建立与选择 5.11

在利用仪器工具收集标准化的定量数据资料过程中，度量技巧起着十分重要的作用。量表和指标是十分重要的，因为它们提供了一些定量的信息，这些量表服务于一个更大的、精确的统计操作和明确的解释。但是在建立一套新的社会量表之前，应进行一个十分仔细的文献检索，以验证已有的研究中的相关量表能否用来测量因变量或自变量。总的原则就是：如果可利用的量表在性质上有效度、信度和效用（在优先权的顺序上），则这个量表就应该被采用。运用这样的量表进行比较研究、纵向研究是可能的。除非变量所代表的因素从未被测量过，否则，千万不要轻易编制一套新的量表。因此，讨论的焦点在于：在找不到合适的量表和研究者决定创建一套指标体系或量表时，如何想出一些测量指标以应用于经验研究。Paul F. Lazarsfeld 和 Morris Rosenberg（1962）对这个问题做了如下回答：

> 第一个步骤似乎是一个非常含糊的形象和结构的创建，这来源于研究者对理论问题各方面细节的全身心投入。这个创造性的行动开始于对一些潜藏的具有共同特征的离散现象的感知，或者研究者可能观察到某些规律并试图对它们做出解释。无论如何，概念首次被创造出来时，它是某个能被模糊觉察到的实体，该实体使得被观察到的事物之间的关系富有意义。接下来的步骤就是用精确的论述来表述这个概念。我们可以应用"方面"、"要素"、"维度"或类似的说明。它们有时在逻辑上来源于一个综合性的概念，或其中一方面通过另一方面演绎而来，或相互之间存在着实验上的相关关系。概念是一组复杂的现象组合而成的，而不是由一个简单的、可直接观察到的项目组合而成的。为了把概念组合成研究设计，必须选择可观测的指标（p. 15）。

“指标”和“量表”这个两术语通常交互使用来指各种各样的测量，绝对关系或相对关系的测量、单一关系或复合关系的测量、简单或精细的测量结果。

指标可能是很简单的。例如，测量道德水平的一种方法就是直接提问：“你认为你的精神状态处于哪个等级？很好、好、一般、差、很差？”这种方法有可能再细分，从而在一个数字化量表上作出反应。注意以下这个量表具有9个点：

很好		好		一般		差		很差
1	2	3	4	5	6	7	8	9

构建量表的基础是逻辑推理，使用数字化量表要求一个心理连续统一体的假设，即被试能够实事求是地进行自我评价。尽管可以进行诸如工作表现、缺席、迟到、喝酒次数和睡眠时间等标准的相关测试，但这样的量表仍需要表面效度。

复合指标是测量的重要组成部分，每个复合指标由简单指标构成，例如精神状态可以被看成多维度的复合指标。以下这些问题可以组成四个测量：

- 你对工作的满意度怎样？
- 你对公司的满意度怎样？
- 你对个人生活满意度怎样？
- 你对所居住的社区满意度怎样？

每个问题的回答都有很好、好、一般、差、很差5个不同选择项，并分别被赋予分值5、4、3、2、1。得分范围从4到20都可能。这样一个复合指标可能会提高精确度、信度和效度。

精确性被给予了比信度和效度更多的关注。在某一点上，一种测量方法有它提升改进的极限，为了提高精确性，有必要使用更精确的方法。许多测量方法把它们自身与线性、等距或等距呈现(equal-appearing intervals)联系在一起，这意味着量表遵循直线模式，其计分系统在交互改变的单元和统计操作基础上得以修订，这是 Thurstone 态度测量技术的一个主要贡献。

单维度或同质性是另外一种想要寻找的特征。单维度或同质性的量表只测量一个维度并无混合因素。这是 Guttman 量表技术首要关注的问题。可再生性(reproducibility)这样一个特征使研究者通过仅知道整个量表分数就能预测被试的回答方式。这一特征根植于加特曼测量技术之中。

情感强度被引入 Likert 量表。通常要求被试在5点量表上(从非常赞同到强烈反对)标注自己的情绪体验。这需要应用项目区分测试(item discrimination)。

没有哪一种单一的方法能囊括所有这些技术的优势。因此，我们要理解这些方法各自的研究目的和它们之间的区别，这是很重要的。

参考文献

Lazarsfeld, Paul F., & Rosenberg, Morris. (Eds.)(1962). The Language of social research: A reader in the methodology of social research. Glencoe, IL: Free Press

5.11.1 Thurstone 等距量表

特性 这种量表由许多项目构成,量表上的这些项目所处的位置由专家通过等级评定操作(ranking operation)所决定。被试选择最能描述他或她感受的选项。

效用 量表接近等距水平的测试,这意味着量表上任何两个数字之间的距离是相等的,可应用于参数统计与非参数统计。有关这方面统计的更多信息参见第6部分。

构建

- 调查者收集了成百上千条确信与调查中的态度有关的陈述。
- 大量专家(50~300)分别把这些陈述分成11组,范围从非常赞同到中立再到强烈反对。
- 根据专家确定的位置把某个陈述的测量值计算为中间值。
- 太离谱的陈述与模棱两可或无关的陈述被舍弃。
- 通过选择从一端到另一端之间均匀分布的项目形成量表。

例子

- Duggan, Ashley, Hess, Brian, Morris, Deanna, Kim, Sooyean, Wilson, Kaeherine(1999, April). 测量学生对因特网教育作用的态度(Measuring students' attitude toward educational use of internet),文章在加拿大 Montreal 召开的美国教育研究协会年会上公开陈述。

调查学生对待互联网的态度的研究设计,是为了编制一个能提供定量测量分析本科大学生对网络教育作用看法的工具。这项研究也调查了与学生态度相关的一些行为。395名本科生对该量表不同形式的回答被用于构建这次测量。对网络教育作用持肯定态度的观点写成两种形式:Thurston 等距量表和 Likert 总加量表。这些试验性的量表与一个社会赞许效应量表同时使用,以确保这些学生不是以一种社会所赞许的方式来回答量表中的题目。根据188个学生的回答,最终形成了有18个项目组成的 Likert 累加量表模式:具有高度内部一致性的"对网络教育作用的态度"(ATEUI)量表。一些行为的相关性增加了量表的结构效度。赞同态度与以下因素有关:①跟踪有价值的教育网点;②与朋友共享网上的信息;③选择使用网络的班级;④在一般意义上和为教育目的而高频率使用网络;⑤使用网络进行教育的许多理由;⑥使用网络的许多特点。结果表明,在对待网络教育作用的态度上,无论在男女性别上,还是在班级之间,均不存在差异。对网络教育作用态度的进一步研究应继续获得一些新的有关行为相关的资料。

- Fukuhara, S., Keller, S. D., Kaasa, J. E., Ware, J. E., Jr., Leplege, A.,

Gandek, B. , Sansom—Fisher, R,W. , Aaronson, N,K. , Sullivan, M. , Alonso, J. , Wood—Dauphinee, S. , Apolone G. , Bjorner, J. B. , Brazier, J. , & Bullinger, M. (1998). 测试被广泛使用的反应选项标签的译文的平衡性:来自IQOLA研究项目的结果(Testing the equivalence of translations of widely used response cholce labels: Results from the IQOLA Project),《临床流行病学期刊》(*Journal of Clinical Epidemiology*),51,pp. 933-944。

SF—36健康调查表上的反应选项标签在意义上的相似性被予以跨国评估。使用Thurston等距量表的变化形式,来自13个国家的专家样本(范围:10到117,中位数:48)对反应选项标签的译文进行等级评定,在表示反应选项标签的重要性(与自己关系亲密的人的健康好坏有密切关系)的一条10厘米直线上,专家标上不同的点作为评估判断。等级评定主要用于决定在反应量表中反应选项标签的顺序上的一致性,相邻选项标签之间的距离程度都是相等的,以及由于选择项、国家、判断标准以及选择项标签和国家之间的相互作用而导致的方差数量。结果证实了反应选项标签的排序。在各国范围内以及所有反应量表中,序数对的百分比范围为88.7%到100%(中数=98 2%)。对反应选项标签的量度的测试,支持了量表"半间距"的特点。方差分析的结果支持反应选项标签量度在不同国家之间的普及性,反应选项标签解释了等级上64%到77%的方差,国家只解释了1%到3%的方差。这些结果都证明了SF—36反应选项标签在国家之间的均等性。在观察中可以看到,背离等距假设的结论在不同国家之间是相似的,在SF—36得分体系下重新校准的两种反应选项量表之间的差异是最大的。结果也证实,在SF—36标准得分下,对各个项目的译文进行评分是正当的。在其他国家中来自同一量表的这些项目是否像在美国那样,都可以从量表假设的测验中获得评估。

- Latter, James M . (1990). Thurstone等距量表问题中的"案例"解决方法:网络信息流通的最小成本法(A minimum-cost network-flow solution to the Case V Thurstone scaling problem),《心理测量学》(*Psycho metrika*), 55, pp. 353-370。

它为解决对配对比较资料中存在的有限不一致性反应不敏感的单维度量表评估提供了一种思路和方法。正如最小成本的网络流量问题所表明的那样,解决的程序体现在与敏感性诊断的结合之中,该敏感性诊断对传统的Thurston等距量表评估中单一配对比较的影响力进行评估。当资料中潜藏偏差时,在保持评估的间距量表特征方面,网络技术似乎比Thurston测量更成功。

- Schriesheim, Chester A. ,& Novelli, Jr. (1989)。对量度评估的间距量表特征的对比测试、"案例Ⅲ"测量以及对等间距频率反应锚点的建议(A comparative test of the interval-scale properties of magnitude estimation and Case Ⅲ scaling and recommendations for equal-interval frequency response anchors),《教育和心理学

测量》(*Educational and psychological measurement*),49,pp. 59-74。

使用量度评估和 Thurston"案例Ⅲ"完全等级的配对比较测试,比较了在所推荐的两套通过测量技术而获得的等间距反应锚点之间的差异。205 个大学生结果中的差异反映了样本中的差异,也反映了任务和计算对数之中的差异。

- Yen,Wendy M. (1986)。教育测量的量表选择:项目反应理论的观点(The choice of scale for educational achievement: An IRT perspective),《教育测量期刊》(*Journal of Educational Measurement*),23,pp. 299-325。

两种建立教育成就等距量表的方法在这里得到讨论:Thurston 的绝对测量法和项目反应理论。选择一个量表的不同标准得以比较。一般认为,要判断所选择的测量程序的合适性和有用性则需要更清晰的标准。

关于 Thurstone 等距量表的读物

Duggan, Ashley, Hess, Brian, Morgan, Deanna, Kim, Sooyean, & Wilson, Katherine. (1999, April). *Measuring students' attitude toward educational use of the Internet.* Paper presented at the annual conference of the American Educational Research Association, Montreal, Canada.

Fukuhara, S., Keller, S. D., Kaasa, J. E., Ware, J. E., Jr., Leplege, A., Gandek, B., Sanson-Fisher, R. W., Aaronson, N. K., Sullivan, M., Alonso, J., Wood-Dauphinee, S., Apolone G., Bjorner, J. B., Brazier, J., & Bullinger, M. (1998). Testing the equivalence of translations of widely used response choice labels: Results from the IQOLA Project. *Journal of Clinical Epidemiology*, 51, 933-944.

Lattin, James M. (1990). A minimun-cost network-flow solution to the Case V Thurstone scaling problem. *Psychometrika*, 55, 353-370.

Schriesheim, Chester A., & Novelli, Luke, Jr. (1989). A comparative test of the interval-scale properties of magnitude estimation and Case Ⅲ scaling and recommendations for equal-interval frequency response anchors. *Educational and Psychological Measurement*, 49, 59-74.

Yen, Wendy M. (1986). The choice of scale for educational measurement: An IRT perspective. *Journal of Educational Measurement*, 23, 299-325.

5.11.2 Likert 量表

特性 这是一个由被试作出反应的一系列项目所组成的一个总加量表。在强度量表上显示出被试对每个项目的支持或反对的意见。这种 Likert 技术生成了一个有序的量表,这个量表一般要求非参数统计。

效用 当这个量表被用于调查人们对某件事情的具有连续性的特定态度或综合态度时,具有很高的信度。分数包括了对每项陈述所表示的强度测量。

构建

- 调查者收集大量的调查项目,这些项目被认为是与所调查的态度有关,并且这

些态度非常明确地表示赞同或反对。

- 将上面这些项目用于一组被试,这些被试代表了该问卷的调查对象。
- 对这些不同项目的反应是按这样的方式来打分的:持最赞成态度的反应就给予最高分。
- 每个被试的总分是通过对他或她在每个项目上的成绩求和计算得出的。
- 分析这些反应,然后看哪些项目在总计分的最高和最低的特征之间具有最显著的差异。
- 区分度最好的项目(至少为6个)用于组成一个量表。

例子

- Belmente-Serrano, M. A., Beltran, Fabregat J., & Paz, Furio M. (1996). 在使用Likert量表研究风湿性关节炎病人的过程中,有关HAQ问卷(20个题项和8个题项的两个版本)的比较研究(A comparative study of HAQ questionnaires—versions of 20 and 8 items—with Likert scale and visual analogue scale in rheumatoid arthritis patients), *Revista Espanola de Reumatologia*, 23, pp. 83-88。

本项研究的目的在于对斯坦福健康评估问卷(HAQ)的残疾量表西班牙语版本与Pincns的8项目缩减版本(MHAQ)作一个交叉确认,并且将形象化类比量表(VAS)和言语顺序Likert量表当作工具来比较,进而评估风湿病人的疼痛和全面健康。给每一个风湿关节炎病人发一张问卷表,这个问卷表包含了HAQ和两个量表的20个题项。作为HAQ本身的一个子集,MHAQ中的8个项目的值可从原始数据中得出。HAQ和MHAQ这两个量表显示出相当好的内部一致性(alpha =9),也显示出好的标准和结构效度。HAQ和MHAQ两者在总分值($r = 0.88, p < 0.001$)方面以及每个子量表方面是有高度的相关性(r值的范围为:$0.57 \sim 0.58, p < 0.001$)。从MHAQ所得到的平均值比从HAQ得到的平均值低31.2%($p < 0.001$)。研究发现,HAQ和MHAQ与Steinbrocker功用性能力量表以及测量疼痛和健康状态的两类量表之间有显著的、中等的、相似的相关性。对疼痛和全面健康量表的研究显示,有关这两个变量的Likert量表和VAS量表有很好的相关性($r = 0.77, r = 0.50, p < 0.001$)。Likert和VAS量表有一种高度的线性关系($p < 0.0001$)。总的来说,可以发现,MHAQ和HAQ一样是合适的。然而,两种问卷之间的平均得分的显著差异使得要笼统地使用它们变得不太可能。对于疼痛和全面健康两种情况,Likert和VAS量表之间所发现的显著相关性和高线性关系表明,使用Likert量表的等价值来代替VAS量表的缺省值是可能的。

- Cheung, K. C., & Mooi, L. C. (1994). 等级量表模型与Likert量表双重测量之比较(A comparison between the rating scale model and dual scaling for Likert scales),《应用心理测量》(*Applied Psychological Measurement*), 18, pp. 1-13。

与间距量表特质的维度以及项目反应功能模式相关的问题可以通过比较测量的方法来解决:项目反应理论模式和双重测量。这两种方法之间的相似性可以通过一项研究得出,这项研究就是让 326 名新加坡 3 年级女大学生去完成一项 Likert 量表。

- Gu, Yongqui, & others(1995),多久才算是经常? Likert 量表在语言学习策略研究中引用的不明确性(How often is often? Reference ambiguities of the Likert-scale in language learning strategy research),《英语语言教学》(Occasional Papers in English Language Teaching),5, pp. 19-35。

这篇文章以个人经验为基础,考察 Likert 五点量表在学习策略启发中的模糊性。让 120 个接受高等教育的非英语专业的中国学生回答 4 张相似的问卷表,每张问卷包含 20 个相同的从 Oxford 量表(1990)中选出的问题项。问卷 1 使用了 Oxford 量表,但却没有规定参考的范围;问卷 2 要求被试先与同年级的同伴进行比较,然后再作选择;问卷 3 要求被试先比较他们过去在中学期间的学习经历,然后挑选出他们现在的行为频率;问卷 4 则要求被试将一个语言行为的出现频率与其他语言技能的出现频率直接比较,然后再核对该语言行为的相关出现频率。调查结果显示,在所使用的 20 个项目中,这 4 个问卷表中有 13 个项目具有显著的差异。文章也讨论了一些问卷研究的方法学应用问题,并且提出了有关进一步研究的建议。

- Hassan, Abdel Moneim Ahmed, & Shrigley, Robert L. (1984),设计一个 Likert 量表用来测量化学态度(Designing a Likert scale to measure chemistry attitudes),《校园科学与数学》(*School science and Mathematics*),84, pp. 659-669。

通过设计一个化学态度量表,提出了一些设计化学课学习态度量表的原则,并且举例说明了化学课学习态度量表设计的程序。利用一个有 20 个题项的化学量表的数据库,对量表的设计程序、效度以及信度逐一进行了讨论。

- Smith-Sebasto, N. J., & D' Costa, Ayres(1995),设计一个 Likert 量表来预测大学本科生由于环境因素而可能表现出来的行为:多步骤过程(Designing a Likert scale to predict environmentally responsible behavior in undergraduate students: A multi-step process),《环境教育》(*Journal of Eviromental Education*),27, pp. 14-20。

本文试图设计一种既可靠又有效的工具来评估强化的控制轨迹和环境性反应行为之间的关系。本文给出了一个六步骤的心理测量程序,这个程序用以为本科大学生编制一套环境性行为内在控制指数(EAICI)。文章包含了 54 个参考书目。

关于 Likert 量表的读物

Belmonte-Serrano, M. A., Beltran, Fabregat J., & Paz, Furio M. (1996). A comparative study of HAQ questionnaires—versions of 20 and 8 items—with Likert scale and visual analogue scale in rheumatoid arthritis patients. *Revista Espanola de Reumatologia*, 23, 83-88.

Cheung, K. C., & Mooi, L. C. (1994). A comparison between the rating scale model and dual scaling for Likert scales. *Applied Psychological Measurement*, 18, 1-13.

Gu, Yongqui, & others. (1995). How often is often? Reference ambiguities of the Likert-scale in language learning strategy research. *Occasional Papers in English Language Teaching*, 5, 19-35.

Hassan, Abdel Moneim Ahmed, & Shrigley, Robert L. (1984). Designing a Likert scale to measure chemistry attitudes. *School Science and Mathematics*, 84, 659-669.

Smith-Sebasto, N. J., & D'Costa, Ayres. (1995). Designing a Likert-type scale to predict environmentally responsible behavior in undergraduate students: A multistep process. *Journal of Environment Education*, 27, 14-20.

5.11.3 Guttman 量表分析

特性 Guttman 量表技术试图确定的是一个量表的单维性。只有那些符合再现性标准的项目才能被接受为具有可测量性。如果量表是单维度的，那么一个人比另一个人有一个更赞成的态度，则他对每一个项目的反应在赞成度方面就应该等于或者大于那个人。

效用 每个分数都与一个非常相似的反应模式或者量表类型相对应。它是少数几个能用分数对所有陈述的反应模式进行预测的量表之一。只需少数几个陈述就可以提供一个可测量的反应范围。我们一起来看一下在表 5.9 中所显示的分析，这个分析显示了 14 个被试如何对几个陈述回答“是”以及分数怎样反映一个特定的反应模式。

表 5.9　Guttman 量表分析中的反应模式

被试	项目 7	项目 5	项目 1	项目 8	项目 2	项目 4	项目 6	项目 3	得分
7	是	是	是	是	是	是	是	—	7
9	是	是	是	是	是	是	是	—	7
1	是	是	是	—	是	是	—	是	6
10	是	是	是	是	是	是	—	—	6
13	是	是	是	是	是	是	—	—	6
3	是	是	是	是	是	—	—	—	5
2	是	是	是	是	—	—	—	—	4
6	是	是	是	是	—	—	—	—	4
8	是	是	是	—	—	是	—	—	4
14	是	是	是	是	—	—	—	—	4
5	是	是	是	—	—	—	—	—	3
4	是	是	—	—	—	—	—	—	2
11	—	—	—	—	是	—	—	—	1
12	是	-	-	-	-	-	-	-	1

构建

- 挑选那些适合应用于可测量目标的陈述。
- 在样本人群中(大约为 100)测试这些陈述。
- 去掉那些超过 80% 同意或不同意的陈述。
- 将被试按照从具有最赞成的态度到具有最不赞成的态度进行排序。从左到右进行排列。
- 把陈述按照从具有最赞成的答案到具有最不赞成的答案进行排序。从左到右进行排列。
- 去掉那些不能区分赞成的被试和不赞成的被试的陈述。
- 计算再生性系数：
  a. 计算错误的数量(那些不适合模式的反应)。
  b. 再生性系数 = 1 -(错误的数量/反应的数量)。
  c. 如果再生性系数等于 0.90 或者更大,那么就可以说有单维量表存在。
- 根据赞成反应的数量或者反应模式,给每个被试打分。

例子

- Burgin, Robert(1989),Guttman 量表分析：应用于图书馆科学(Guttman scale analysis: An application to library science),《图书馆和信息科学研究》(*Library and Information Science Research*),11,pp. 47-57。

对 Guttman 量表分析技术进行概述，并且大致描写它在社会科学研究中的作用。为了阐明该量表在图书馆科学中的潜在应用，编制了一种处理过量书本的严格的 Guttman 量表，1986 年对公共图书馆进行调查而获得的数据被写进这个量表。

- Kramer, Deirdre A. (1983)，相对思维和辩证思维的初步调查研究（A developmental investigation of relativistic and dialectical thought），此文在旧金山举行的老年学研究学会年度科学会上报告。

后形式运算思维具有相对主义和辩证主义的特点。为了检验成人的相对思维和辩证思维之间的年龄差异，以及考察形式运算对于这些思维形式是否是必需而非充分的，需要对三组成年人：20 个年轻人（平均年龄为 19.6 岁），20 个中年人（平均年龄为 46.2 岁）和 20 个老年人（平均年龄为 68.5 岁）实施三个认知任务。Ammons 快速测试被用于考察有无言语智力的差异。接下来，对被试进行四组形式运算任务——变量分离，三种对两种框架体系的协调措施的测量，以及两个生活中的两难问题要求被试作出反应。将对进退两难问题作出的反应分为四种思维类型（形式—机械主义，相对主义，自相矛盾的意识，自相矛盾和辩证主义合二为一）。分析结果显示，在快速测试中老年人组比年轻人组得分高得多，而中年成人组得分介于前面两者之间。在形式运算任务中，这几组成人的表现都很完整。在面对像生活中一样的进退两难问题时，老年成人组表示出更少的抵制，显示出对相对和辩证思维更大的接受性。Guttman 量表分析表明，对于辩证思维形式运算是必需但非充分的。分析结果为"辩证思维是后形式运算的"这一假设提供了潜在的支持。

- Lange, A, Kooiman, K., Huberts, L., & Van Oostendorp, E. (1995)，精神病人精神病理学程度同儿童时期有害性行为的关系：使用新的追忆问卷研究（CHUSE 问卷）（Childhood unwanted sexual events and degree of psychopathology of psychiatric patients: Research with a new anamnestic questionnaire），《斯堪的那维亚精神病学报》（*Acta Psychiatrica Scandinavica*），92，pp. 441-446。

通过一套最近构建起来的回忆性测试工具，即儿童有害性活动（CHUSE）问卷调查表，在 152 名女性精神病患者当中，研究者调查了她们儿童时期受到性威胁或者性虐待的发生率。研究者描写了该问卷的构建和适用性情况。一项 Guttman 量表分析的结果表明儿童有害性活动（CHUSE）具有单维的结构（性虐待的严重程度）。在这些精神病患者当中，受过性虐待的女性明显比那些未受过性虐待的女性更具有一些精神病症状。同时，研究者也对虐待的严重程度和精神病理症状的严重程度之间的相互关系进行了调查。研究者也将关于性虐待问卷调查表的使用情况和其他一些更常见的会谈技巧进行了比较。最后还给出了一些作进一步研究的建议。

- Stempel, Guiod H., Ⅲ. (1982), 柏格尔法院新闻案件判决的 Guttman 量表分析 (A Guttman scale analysis of the Burger Court's press decisions), 《新闻季刊》(*Journalism Quarterly*), 59, pp. 256-259。

文章分析了最高法院中对 47 件与新闻出版有关案子进行表决的情况。分析显示:这些表决情况组成了一个单维的可测量的体系。作者认为这个维度就是政治倾向。

- Von Korff, M., Ormel, J., Keefe, F. J., & Dworkin, S. F. (1992), 对慢性疼痛的严重性进行分级(Grading the severity of chronic pain), 《疼痛》(*Pain*), 50, pp. 133-149。

这个研究设计并评价了一种对慢性疼痛程度进行评分的简单方法,这种方法可用于对普通人口疼痛情况调查,也可用于疼痛病人基本治疗的研究当中。病人疼痛的强度、障碍(disability)、疼痛的持续时间,以及最近疼痛发作情况得到了测量,这一测量在对背痛(测量人数为 1 213 人)、头痛(人数为 779 人),以及下颌功能紊乱疼痛(人数为 397)病人基本治疗情况的一个纵向研究当中进行,其目的是对慢性疼痛程度进行等级评定。Guttman 量表分析表明,疼痛强度以及障碍的测量结果形成了一个可信的等级性量表。疼痛强度测量结果似乎衡量出综合严重程度的更低范围,而障碍的测量结果则标示出综合严重程度的更高范围,最近疼痛发作情况以及近 6 个月疼痛发作的天数的测量并不与疼痛强度或者障碍一起进行。采用简单的打分规则,疼痛强度可以被分成四个等级:等级一,低障碍—低疼痛强度;等级二,低障碍—高疼痛强度;等级三,高障碍—中等限制;等级四,高障碍—严格的限制。对于每个疼痛部位,在基准线被测量的慢性疼痛等级在数据统计上显示出相当的显著性,并且这些等级与下列一些情况之间存在着一种单调的增长关系,这些情况就是:失业率,与疼痛相关的功能性限制,抑郁,相当差的自我健康评价,经常使用鸦片制剂止痛,在基准线以及此以后的一年时间里经常由于疼痛去看医生。疼痛的天数与以上这些变量也有关系,但不如慢性疼痛等级与这些变量的关系明显。与那些稍前一些时间的例子相比,最近疼痛发作的例子(第一次发作在头三个月内),在心理和举止行为功能紊乱方面并未有什么不同。根据以患者为基础的纵向研究数据,慢性疼痛等级可以预测疼痛在开始两周的发生情况,并且这些等级也可以预测接下来三年后与疼痛相关的一些功能限制。如果想要对综合性疼痛程度进行简要而有序的测量的话,对慢性疼痛进行分等级并把其作为疼痛的强度和与疼痛相关的障碍的一个函数,这种做法是有效的。在一段固定的时期内以疼痛发生的天数测量出来的疼痛持续期,也能提供有用的信息。

关于 Guttman 量表的推荐读物

Burgin, Robert. (1989). Guttman scale analysis: An application to library science. *Library and Information Science Research*, 11, 47-57.

Gordon, Raymond L. (1977). *Unidimensional scaling of social variables.* Riverside, NJ: Free Press.

Kramer, Deirdre A. (1983, November). *A developmental investigation of relativistic and dialectical thought.* Paper presented at the Annual Scientific Meeting of the Gerontological Society, San Francisco.

Lange, A., Kooiman, K., Huberts, L., & Van Oostendorp, E. (1995). Childhood unwanted sexual events and degree of psychopathology of psychiatric patients: Research with a new anamnestic questionnaire (the CHUSE). *Acta Psychiatrica Scandinavica*, 92, 441-446.

Lin, Nan. (1976). *Foundations of social research.* New York: McGraw-Hill.

Stempel, Guiod H., Ⅲ. (1982). A Guttman scale analysis of the Burger Court's press decisions. *Journalism Quarterly*, 59, 256-259.

Von Korff, M., Ormel, J., Keefe, F. J., & Dworkin, S. F. (1992). Grading the severity of chronic pain. *Pain*, 50, 133-149.

5.11.4 量表区分技术(scale discrimination technique)

特点 这种技巧试图编制一组项目,这些项目既符合单维测量的要求,也拥有均等出现的间距,并且能够测量强度。这种由 Edwards 和 Kilpatrick 研发出来的技术,结合了 Thurstone 等间距、Likert 总加量表以及 Guttman 量表分析这些建构的一些方面。

效用 结合了以上三个测量技术的各自不同的优势。Thurstone 技术的间距测量质量能得到保证。被试与增加强度测量之间的可区分性则得益于 Likert 量表,测量的单维性则源于 Guttman 测量技术,要注意的是,项目分析会删除量表中间的一些项目。

构建

- 研究者挑选大量被认为可以应用于所测量的态度的一些表述。
- 剔除那些含糊或太极端的题项。
- 把这些陈述交给专家,让他们对每个陈述的赞成程度作出评价,并且将其放入 11 个范畴之中的一个当中。
- 剔除一半具有最大分散性或分歧的题项。
- 赋予剩余题项的分值,作为专家所给分值的中数。
- 将那些陈述改编成总加量表的形式并且对其进行一轮新的评判。
- 进行项目分析以决定哪些问题能够在最低和最高属性之间有最好的区分度。
- 选出最终量表所需项目两倍数量的项目。从每个量表间距之间,选出那些区分度最好的陈述。
- 将这些陈述分为两半,并分别用于两个测试组。
- 为两个测试组确定再现性系数,那些系数在 0.90 或以上的被使用。

例子

- Crist. D. A., Rickard, H. C., Prentice, Dunn S., & Barker, H. R. (1989), 放松训练目录:放松训练效果的自我报告量表(The relaxation inventory: Self-report

scales of relaxation training effects),《个性评估期刊》(*Journal of Personality Assessment*),53,pp. 716-726。

文章对一种用于评价放松训练作用的自我陈述测量法的设计情况进行了研究。研究采用了一种严格的量表构建的统计方法,这一测量方法由修改后的量表区分技术组成,这样就产生了一个有45个题项的问卷表,这个问卷表代表了三个互不相关的量表。这三个量表,即生理焦虑程度、身体评价以及认知紧张状态,与KR20的信度系数分别为0.89、0.95、0.81,显示出高度的内在一致性。在对预测性效度的再次研究中,40位个体被随机地赋予以下一个实验情境:放松训练,紧张情况诱导,前控制,后控制。变异的单变量分析表明,在目录的三个维度中每一个都具有有意义的发现。生理紧张状态量表发现,在紧张情况诱导之后出现生理紧张状态的显著增长;而身体评价量表和认知紧张状态量表也发现,在放松训练之后,放松效果有增加。在目录中,文章也给出了一些进一步研究的建议。

- Foster, Don. (1991),南非社会心理学(Social psychology in South Africa), Johannesburg, South Africa: Lexicon。

大部分态度测量的基本方法都是20世纪二三十年代在美国发展起来的。这些测量方法包括社会距离量表,Thurston等距量表,Likert测量量表,卡兹和布雷的原型评价法。测量区分技术则是在20世纪40年代开发出来的。紧跟在这个技术之后,另一个流行的测量法,语义微分法又开发出来了。这些方法中的大部分都在南非得到应用,并且只要应用得上,这些方法都以南非人为被试逐一加以讨论。

- Killian, Kieran, Watson, Richard, Otis, Joceline St. Amand, Timothy A., & O'Byrne, Paul M. (2000),支气管炎剧烈发作时的症状知觉(Symptom perception during acute bronchoconstriction),《美国呼吸系统疾病和紧急护理学》(*American Journal of Respiratory and Critical Care Medicine*),162,pp. 490-496。

这项研究存在着一个假设:急性支气管炎缩小症所表现出来的症状强度上的某些可变性,可能是由几种刺激变化的强度所造成的,同时产生了一些可通过具体的、可描写的症状来确认的知觉。总共有232个被试吸入醋甲,这些醋甲是双倍浓度,吸入量比最大吸入量少20%,或者减少64毫克/毫升。这一研究确定了呼吸困难的流行情况、与呼吸行为有关的不太具体的不适情况,以及10种具体症状的表述。每一症状的强度通过Borg量表单位作出评价。具体症状所引起的呼吸困难强度的情况可由下面这一方程式($r=0.84$)表示:呼吸困难 $=0.44+0.19$(有困难的呼吸)$+0.41$(胸闷)$+0.20$(无呼吸情况)$+0.14$(吃力呼吸)$+0.11$(胸痛)。呼吸困难的强度增强与下列情况有关:支气管狭小,最初的肺部损伤,体重以及性别(女性)。呼吸困难强度的减少与年龄(更老的年龄)有关;并且当气管对醋甲的反应性增强时($p<0.05$ 适用于氛的

因素),这种情况也会发生。胸闷和胸痛处在区别量表的边缘地带(如,很容易区分出)。胸闷、有困难的呼吸、吃力这几种情况却不太容易区分。

- Velozo, C. A. , Magalhaes, L. C. , Pan, A. W. , & Leiter, P. (1995),功能量表在入院和出院方面的辨别力:康复水平量表Ⅲ(LORS-Ⅲ)的 Rasch 分析(Functional scale discrimination at admission and discharge: Rasch analysis of the Level of Rehabilitation Scale-Ⅲ),《生理医学和康复档案》(*Archives of physical Medicine and Rehabilitation*),76,pp. 705-712。

本项研究的目的在于确认康复水平量表Ⅲ(LORS-Ⅲ)的结构效度,特别关注的是该工具对康复住院病人的日常生活活动/走动情况和病人入院及出院时的交际/认识能力进行区分的能力。文章对 LORS-Ⅲ美国数据系统(LADS)现存的数据进行了 Rasch 分析,该数据采自于 3 056 名住院康复病人(这些病人有肌骨骼受伤的,脑血管突发的有多处受伤/多重疾病的,脑受伤的,神经性肌肉紊乱的,脊髓受伤的),这些数据是在 1992 年 4 月到 1993 年 1 月这段期间进入 LADS 的。LORS-Ⅲ包括 17 个测量区域,这些区域代表了 ADL 中的能力:运动、交际、认识以及记忆方面的能力。这些测量区域当中的 14 个区域由护士和指定的康复病治疗专家来同时打分,这样总共就有 31 个项目。该分析与其他功能性状态测量所报道的发现是一致的。这一分析表明,LORS-Ⅲ包括两个单维性量表,一个 ADL/运动量表以及一个交际/认识量表。尽管所有量表都符合 Rasch 测量模式,但是入院时的 ADL/运动量表用于样本的能力水平目标是最合适的。在出院时,ADL 量表总的来说显得太简单了,因为样本能力水平在朝着功能独立性向上提升。在入院和出院时的交际/认识量表都显示出一种相似的"天花板"效应。这些结果表明,决定功能性状态的测量质量在入院和出院评价中都很重要。像 Rasch 这样的分析为工具完善提供了逻辑方向。

关于量表区分技术的推荐读物

Crist, Dwayne A. , Rickard, Henry C. , Prentice, Dunn Steven, & Barker, Harry R. (1989). The Relaxation Inventory: Self-report scales of relaxation training effects. *Journal of Personality Assessment*, 53, 716-726.

Foster, Don. (1991). *Social psychology in South Africa.* Johannesburg, South Africa: Lexicon.

Killian, Kieran, Watson, Richard, Otis, Joceline, St. Amand, Timothy A. , & O' Byrne, Paul M. (2000). Symptom perception during acute bronchoconstriction. *American Journal of Respiratory and Critical Care Medicine*, 162, 490-496.

Velozo, C. A. , Magalhaes, L, C. , Pan, A. W. , & Leiter, P. (1995). Functional scale discrimination at admission and discharge: Rasch analysis of the Level of Rehabilitation Scale-Ⅲ. *Archives of Physical Medicine and Rehabilitation*, 76, 705-712.

5.11.5　等级评定量表(rating scales)

特性　这个技术以个人判断为基础,试图获得一种对个人、群体或机构特征作出的评价或量化评判。评价者将被评价的个人或对象放在一个连续体的某个点上,或放于一系列有序范畴当中,每个点上或者每个范畴都被赋予一个数值。

效用　等级评定量表可被用于评价态度、价值观、标准、社会活动和社会结构特征。

构建

- 将被测量的连续体分成一些最佳的量表分区数字(大约是5~7)
- 连续体应该没有断续或者分隔的地方。
- 正向的和反向的端点应该交替出现。
- 每个特质由一个问题体现,评价者对这个问题给出一个答案。
- 用一些描述性的形容词和短语来定义连续体中不同的点。
- 调查者应该事先决定某一特质可能的极端情况。这些极端情况可能会在量表要测量的那个群体中出现。
- 只有那些普遍能被人理解的描述性术语才应该被使用。
- 最后的表述不应该在意义上太极端,这样评价者才会避免极端。
- 描述性的短语需要均匀地放置。
- 在试测时,如果回答者对评价有疑问,或者对连续体上不同的点他们有不清楚的地方,研究者都要求他们提出来。
- 被赋予的数值被用于计分。

例子

- Linacre, John M.(1999),对等级评定量表类别效用的调查(Inuestigating rating scale category utility),《效果测量期刊》(*Journal of Outcome Measurement*),3,pp. 103-122。

文章提到8条指导原则,这些指导原则可帮助分析者调查等级评定量表类别是否可以组合产生一些观察资料,从而根据这些观察资料形成一些有效的测量。在Rasch分析环境下讨论这些指导方针,并阐述它们的作用。

- Varner, Roy V., Chen, Y. Richard, Swann, Alan C., & Moeller, Frederick G.(2000),严重病人治疗效果测量工具——简明精神病等级评定量表:探索性研究(The brief psychiatric Rating Scale as an acute inpatient outcome measurement tool: A pilot study),《临床精神病学期刊》(*Journal of Clinical Psychiatry*),61, pp. 418-421。

由于呆在精神病医院的时间长短可能会对病人在出院时的结果产生让人不能接

受的影响，因而就需要使用一个有效的测量工具来评价病人在短期的住院期间（一般为7天）出现的一些有意义的变化，这一工具也用来为不理想的短期结果提供早期的预测。本项研究考察简明精神病治疗评价量表（BPRS）作为一种测量工具的效用。在为期两个月的测试期间，先后有87位成年人进行了BPRS测试。这些病人住在急性精神病病房内，在住院后2天、7天、14天后对之进行测试直到他们出院。在数据分析中使用了总量表BPRS分数和四个分量表分数，并且分析中还包括配对的 t 测试以及相关分析。BPRS总量表分数的平均值显示出在2天、7天以及14天的住院时间里病人有显著的改善（$p<0.001$）。分量表分数以及这些分数与最终结果的关系会随着治病群组的不同而变化。因此，BPRS似乎是一种很有效的病人住院结果测量方法，因为这一方法能够显示出在一周或更少的时间内治疗显著性的变化。分量表的分数可能能够提供更具体的对变化的预测情况，可能有助于分清每个病人的结果，而这些病人在总分数上显示不出显著的变化。

- Ward, M. M., Marx, A. S., & Barry, N. N. (2000)，在全身性红斑狼疮（SLE）中把等级评定量表偏好测量作为一个评估性测量（The rating scale preference measures as an evaluative measure in systemic lupus erythematosus），《狼疮》（*Lupus*），9，pp. 696-701。

偏爱测试可能是有用的测试工具，它们可被用于评价病人全面的与健康有关的生活质量情况。在对23个病人的临床疾病活动和症状变化的预期纵向观察研究中，研究者们研究了等级评定量表偏好测量在全身性红斑狼疮（SLE）病人中的效度和对变化的敏感性，并且将其特征与SLE活动的病人全面评价特点进行了比较。病人每2周评价一次，一共评价40个星期。结构效度可通过在一些变化中的相互关系的强度来作出评价。这些变化包括等级评定量表偏爱测试和病人全面评价中的变化，以及在医生全面评价：全身红斑狼疮测量（SILAG）、欧洲狼疮情况普查（ECLAM）、英国岛屿狼疮评价群体指数（BILAG）和全身性红斑狼疮疾病活动指数（SLEDAI）之中的变化。比起在病人全面评价中的变化，等级评定量表中的变化与这些标准中的每一个的变化有更一致的关系，同时表现出这种测量的结构效度。通过使用医生全身评价或者以SLE活动测量中最大变化的2周时间间隔作为标准，可以测量出对变化的敏感性。当以SLAM或是ECLAM作标准时，对等级评定量表变化的敏感性比对病人全面评价变化的敏感性的一半还要少。虽然这些结果可以为等级评定量表作为一种测量与健康有关的生活品质的方法的效度提供支持，但其对变化作出的有限的敏感性使其作为在临床试验中的一种终端测量方法变得不太有吸引力。

- Wright, Benjamin D., & Masters, Geofferey N. (1982)，《等级评定量表分析》（*Rating scale analysis*），芝加哥：MESA。

这本书讨论了变量的构建和测量方法的设计过程。该书一开始就概述了一个数

字在具有作为某种事物的测量资格之前所应具备的性质。其理论基础是 G. Rasch 的测量哲学。构成良好测量的第一要求就是良好的素材。为了达到比较的可能性,数据必须包含单一变量的可能性,根据这一变量可以对不同的人进行测量。第 2 章给出了一系列必须被考察的数据,并且还评述了一些用于考察数据的技巧。第 3 章描述了 5 种不同的测量模式,每一模式都是为每一种特殊的数据开发出来的。在测量方法的文献中还有一些其他的模式,但是这些都是单个类别的成员,它们得到使用是因为它们符合测量的标准要求。第 4 章探讨的是怎样使用这些模式来获得结果,并且还描述了四种不同的评估程序:PROX,RAIR,UCON 以及 CON。在第 5 章中则讨论了变量的质量控制问题。接下来,第 6、7、8 章列举了所讨论过的技巧的应用问题,并且还使用了四种不同的数据,这些数据被收集起来用于测量药品使用、对犯罪的畏惧、基本的物理学知识以及儿童发展情况这些问题。

关于等级评定量表的推荐读物

Linacre, John M. (1999). Investigating rating scale category utility. *Journal of Outcome Measurement*, 3, 103-122.

Varner, Roy V., Chen, Y. Richard, Swann, Alan C., & Moeller, Frederick G. (2000). The Brief Psychiatric Rating Scale as an acute inpatient outcome measurement tool: A pilot study. *Journal of Clinical Psychiatry*, 61, 418-421.

Ward, M. M., Marx, A. S., & Barry, N. N. (2000). The rating scale preference measure as an evaluative measure in systemic lupus erythematosus. *Lupus*, 9, 696-701.

Wright, Benjamin D., & Masters, Geofferey N. (1982). *Rating scale analysis.* Chicago: MESA.

5.11.6 潜在距离量表(latent distance scales)

特性 这种量表的分析是以一种概率模型为基础的,通过提供序列化信息,这个模型试图将因素分析的一些规则运用到定性数据上。基本的假设是:存在着一套潜在的等级,以至问卷上的任何两个项目或多个项目之间的外显关系都可以通过这些潜在的等级的存在以及这些等级本身来得以解释。

效用 通过对几个类属的外显(观察到的)变量中的关系结构的分析,潜在等级分析提供了一个有关类属的潜在(没有观察到的)变量的描述。这一方法通常被称为类似于因素分析的聚类分析。

构建

- 调查者将被认为与潜在态度相关的问题列出来。
- 问题的答案按照"肯定的—否定的"、"赞成的—不赞成的"等这样的方式进行二分。
- 算出那些每个反应中显示出潜在态度的被试的比例。
- 以其明显的标记来安排项目。
- 潜在等级的频率通过反向概率程序被计算出来。
- 反应模式根据平均的潜在位置来分级,或者使用一个指标来概括每个反应模式

的特点。

例子

- Eshima, Nobuoki (1991), 潜在量图分析(latent scalogram analysis),《行为测量学》(*Behaviormetika*), 20, pp. 1-21。

在科学研究中,反应结构经常要比那些在潜在距离模式下假定的结构更复杂。这一潜在距离模式其实是 Guttman(1950)所提出的量图分析的扩展。文章对测量模式进行了评论和比较,然后扩展到心理测验量图分析的设计。在对线性的和分支的等级结构进行提取时,在探索性和验证性环境下使用了模型选择程序。文章从数学的角度提供了对潜在量表的动态解释,并且还提出了对潜在量表的比例进行解释的方法。数字分析显示出这一方法的有效性,这一方法用于在二分数据的基础上得出一个简单的潜在结构,并且还可解释所提取的结构。

- Eshima, Nobuoki, & Asano, Chooichiro. (1998), 论潜在距离分析和 MLE 运算法则(On latent distance analysis and the MLE algorithm),《行为测量学》(*Behaviormetrika*), 24, pp. 25-32。

文章首先讨论了 Lazarsfeld 和 Henry 提出的潜在距离分析,在这一分析中所有的二分项目(是/否)都是由一个普通的因素来决定的,并且根据潜在因子的水平,样本中的每一个体都与每一项目在反应可能性上存在着一致。文章还提出了一个运算法则,在将潜在反应参数转换成对数形式以及运用最新的最大似然估计(MLE)方法之后,这个运算法则会给出合适的解决方法。文章也提供了实际的数据分析。

- Mellenbergh, Gideon J. (1994), 连续项目反应的一维潜在特质模型(A unidimensional latent trait model for continuous item responses),《多元行为研究》(*Multivariate Behavioral Research*), 29, pp. 223-236。

文章描述了一套用于连续项目反应的一般线性潜在特质模型。这个用于连续项目反应的特殊线性例子是 K. G. Joreskog 的同类项目反应模型。用于连续和二分的项目反应模式之间的一致关系比平常的假设更紧密。

- Pascual, Leone Juan, & Baillargeon, Raymond. (1994), 心理注意的初步测量(Developmental measurement of mental attention),《行为发展国际期刊》(*International Journal of Behavioral Development*), 17, pp. 161-200。

文章展示了一个心理注意和工作记忆的辩证的构建主义模型,这一模型被用来解释研究参与者在误导性测试项目中的加工过程。从语义上导出的一组 10 个理论的结

构性预测揭示了心理注意资源和项目不同的心理需要之间的关系，这些不同的项目共同决定可能的行为表现。文章对这些预测进行了评价。研究测试了一组总数为616的儿童(年龄为5~14岁)。测试结果显示：①数据符合Lazarsfeld的潜在距离模式，测试还为前面的10个预测提供了支持；②当从行为上来对儿童进行评价时，儿童的心理能力(潜在心理能力类别)可能会随着年龄增长以一种不连续的方式增长，并且这一能力有生成间距量表的可能性；③统计者经常认为的"测量误差"似乎(部分的)是信号上的而非噪音上的。

- Windle, Michael, & Dumenci, Leyent. (1998)，用潜在特征状态模型对母亲和青春期的抑郁情绪调查研究(An investigation of maternal and adolescent depressed mood using a latent trait-state model)，《青春期研究期刊》(*Journal of Research on Adolescence*)，8，pp. 461-484。

有一项研究使用潜在特征状态模型来研究母亲和青少年的抑郁情绪。这一研究发现，母亲的抑郁可以通过较低的家庭收入、较低的家庭内聚力、较少感觉到社会支持以及较高的父母角色压力这些情况来得到预测。研究还发现，青少年抑郁特质则可以通过较少感觉到家庭支持、更低的平均学习成绩、较多有压力的生活事件以及女性性别这些因素来预测。

关于潜在距离量表的推荐读物

Eshima, Nobuoki. (1991). Latent scalogram analysis. *Behaviormetrika*, 20, 1-21.

Eshima, Nobuoki, & Asano, Chooichiro. (1988). On latent distance analysis and the MLE algorithm. *Behaviormetrika*, 24, 25-32.

Guttman, L. (1950). The basis for scalogram analysis. In S. A. Stouffer with the Social Science Research Council(Eds.), *Studies in social psychology in World War II: Vol. 4. Measurement and prediction.* Princeton, NJ: Princeton University Press.

Langeheine, Rolf, & Rost, Jürgen. (Eds.). (1988). *Latent trait and latent class models.* New York: Plenum.

Lazarsfeld, Paul F., & Henry, Neil W. (1968). *Latent structure analysis.* Boston: Houghton Mifflin.

Mellenbergh, Gideon J. (1994). A unidimensional latent trait model for continuous item responses. *Multivariate Behavioral Research*, 29, 223-236.

Pascual, Leone Juan, & Baillargeon, Raymond. (1994). Developmental measurement of mental attention. *International Journal of Behavioral Development*, 17, 161-200.

Windle, Michael, & Dumenci, Leyent. (1998). An investigation of maternal and adolescent depressed mood using a latent trait-state model. *Journal of Research on Adolescence*, 8, 461-484.

5.11.7 配对比较法(pared comparisons)

特性 这一技巧试图在不了解任何相对应的被试数值的情况下测定定性刺激的心理值。要求被试在一组配对的陈述或对象当中选出最赞成的一对，文章试图对陈述

和对象在一个连续体上进行排序。这一方法也被称作迫选法。

效用 通过配对比较进行排序,对于保证在连续体上有一个精确而且相对的位置来说是一种相对较快的过程。在随意的等级评定方法中,对比性排序通常能够增加信度和效度。

构建

- 调查者选出那些与要测量的特征有关的陈述。
- 陈述按照下面所示的所有可能的组合方式来排列:组合的数量为 $N(N-1)/2$,这里 N 为陈述的数量。
- 要求被试选出每一对中的哪个陈述是更赞成的。
- 计算出每一陈述相对于每个其他陈述所具有的判断比例。
- 再计算每一陈述的比例总和。
- 把这些比例再转化成标准测量值。
- 通过计算绝对平均差异来检验内在一致性。
- 向被试呈示陈述,要求他们标出对每个陈述的赞成和不赞成程度。
- 每个被试的分值就是他或她的赞成反应的中值。

例子

- Eisenherg, L. S. , & Dirks, D. D. (1995),儿童配对比较法和类别等级评定的信度和敏感性(Reliability and sensitivity of paired comparisons and category rating in children),《听说研究期刊》(*Journal of Speech and Hearing Research*),38, pp. 1 157-1 167。

在这项调查中,研究者评价了孩童对言语清晰的主观判断情况,判断采用了配对比较和类别等级评定的方法。80 个年龄为 4 ~ 8 岁的、有正常听力的孩童来判断句子的清晰度,通过使用一些增加儿童理解力的条件使这些句子被系统地作了带通过滤(bandpass-filter),儿童的理解力可通过发音清晰指数估计出来。被试被分成 4 个年龄组(4 岁、5 岁、6 岁,以及 7 ~ 8 岁组),每组 20 人,由于使用一些适合于孩童的素材和训练方法,通过两个心理程序(每个程序用于每个年龄组当中的 10 个被试),就可得出判断。结果显示,无论使用哪一个程序,5 岁年龄组或更大年龄组的小孩能够作出可靠的清晰度判断。然而,比起类别等级评定法,配对比较法在探测带通过滤条件之间的差异方面更具敏感性。

- Eisenherg, Laurie S. , Dirks, Donald D. , & Gornbein, Jeffrey A. (1997),使用配对比较法和类别等级评定法测量言语清晰程度的主观判断(Subjective judgments of speech clarity measured by paired comparisons and category rating),《耳朵与听力》(*Ear and Hearing*),18, pp. 294-306。

这项研究的目的在于根据配对比较和类别等级评定来比较听者对言语清晰度的

主观判断。这项研究采用了一些刺激条件,这些刺激条件在刺激项目之间以相对的距离发生变化,从而产生大范围的行为表现,或产生小范围的行为表现。研究测量了12个有正常听力的成人(实验一)和8个听力受到损害的成人(实验二)对言语清晰度的主观判断。句子通过6个带通滤波器得到处理。这些滤波器可以单纯地增加发音指数(AI)估计值。这些句子就构成了将被判断的刺激物。使用6个滤波器中的3个,可以另外再获得一些主观性的判断。这些判断可以用作刺激条件,在刺激条件中行为表现的范围很广(AI中的较大差别)也很窄(AI中的小差别)。结果表明,由配对比较法和类别等级评定法测量得出的言语清晰度判断对于有正常听力的被试和听力受损的被试与AI评估值都有较高的相关性。对于听力受损的被试来说,表明比较要比类别等级评定更具敏感性。当行为表现范围很窄时,对于两组被试来讲,表明比较要比类别等级评定在区分不同滤波器中更具敏感性。

当配对比较数据被转化成一个可与类别等级评定相比较的量表时,这种差异性对于听力正常的被试来讲就不太明显。在心理测量程序当中,被试之间较大的变化对于听力受损的被试来讲则很明显,对于类别等级评定来说则更值得注意。研究者总结如下:当判断行为表现在一个较大范围内变化的各种刺激项目之间的清晰度时,类别等级评定法和配对比较法都为听力正常的被试提供了可比较的判断。在一些情况中,其刺激项目之间的知觉差异要么受由条件的选择所限制,要么是受感觉神经的听力损失的影响所限制。对于这些情况,配对比较方法更敏感。

关于配对比较法的推荐读物

Bonebright, T. L. (1996). An investigation of data collection methods for auditory stimuli: Paired comparisons versus a computer sorting task. *Behavior Research Methods, Instruments, Computers*, 28, 275-278.

Bossuyt, P. (1990). *A comparison of probabilistic unfolding theories for paired comparisons data.* New York: Springer Verlag.

David, H. A. (1963). *The method of paired comparisons.* Port Jervis, NY: Lubrecht and Cramer.

Eisenberg, L. S., & Dirks, D. D. (1995). Reliability and sensitivity of paired comparisons and category rating in children. *Journal of Speech and Hearing Research*, 38, 1157-1167.

Eisenberg, Laurie S., Dirks, Donald D., & Gornbein, Jeffrey A. (1997). Subjective judgments of speech clarity measured by paired comparisons and category rating. *Ear and Hearing*, 18, 294-306.

5.11.8 语义微分法(semantic differential)

特性 语义微分法试图测量出一个物体对个体的意义。要求被试对一些给定的概念(如,"非裔美国人","共和国","妻子","我想要成为的我","像我一样的我")在一个7点两极等级评定量表上进行打分评价。任何一个概念都可以被评分,不管是一个政治问题、一个人、一个机构,还是一件艺术品。这些7点量表包括像下面一样的两极量表:①公平—不公平,干净—肮脏,好—坏,有价值的—无价值的;②大—小,强—弱,重—轻;③积极—消极,快—慢,热—冷(如表5.10所示)。根据被试对这个

形容词与这个词或者概念的相关性或者联想的感觉程度来评分。测量了三个亚群体被试态度的如下三个维度:①个体对要评分的物体或概念进行评价,评价要与更传统的态度量表中的赞成—不赞成维度相对应;②个体对物体或者概念的力量或者能量的感知情况;③个体对物体或者概念活动的感知情况。

效用 在10~15分钟的时间里可以进行一个含有100个项目的测试。而一个含有400个项目的测试则需要一个小时,语义微分法通过对概念和量表的选择,可用于对许多现象的研究当中。这一方法在构建和分析社会测量量表时也是很有用的。

构建

- 调查者准备一些概念,这些概念适合指导所测量的变量的理论。
- 根据以前的理论基础,选择成对的两极形容词。
- 在进行一项形容词等级评定工作时,通过要求不同的小组(对比或实验控制设计),作出规定的定位从而与等级评定任务相对应,这样就可以实验性地对形容词作出选择。比如:当一组被试认为,一个人若持有肯定的态度就会对概念进行等级评定时,他们就会被要求进行等级评定。而其他的被试认为,如果一个人持强烈的否定态度就会对概念进行等级评定时,他们也会被要求进行等级评定。为了使用语义微分的形式,可以让被试获得一些标准指导。研究者分析数据后,挑选一些可以明确区分小组的成对形容词。
- 选出新的几组被试,这些被试在给概念评定等级时可以作出规定的定位,并分析数据。

表5.10 语义微分量表的例子

15个概念:爱、孩子、我的医生、我、我的工作、精神病、我的母亲、内心的宁静、骗子、我的配偶、自我控制、仇恨、我的父亲、混乱、性。每个概念在下列10个量表中进行等级评定。

有价值的	___:	___:	___:	___:	___:	___:	___:	无价值的
清洁的	___:	___:	___:	___:	___:	___:	___:	肮脏的
美味的	___:	___:	___:	___:	___:	___:	___:	味道不佳的
大的	___:	___:	___:	___:	___:	___:	___:	小的
强的	___:	___:	___:	___:	___:	___:	___:	弱的
深的	___:	___:	___:	___:	___:	___:	___:	浅的
快的	___:	___:	___:	___:	___:	___:	___:	慢的
主动的	___:	___:	___:	___:	___:	___:	___:	被动的
热的	___:	___:	___:	___:	___:	___:	___:	冷的
紧张的	___:	___:	___:	___:	___:	___:	___:	放松的

资料来源:Chales E. Osgood and Zalla Luria (1954),"A Blind Analysis of A Case of Multiple Personality Using the semantic Differential." *Journal of Abnormal and Social Psychology*,49,579-591.

例子

- Bishop,J. Joe.(1999,April),对捷克的民主主义的定位:三种不同类型的中学生和中学教师对民主主义含义的语义微分分析(Locating Czech democracy: A se-

mantic differential analysis of the meaning of democracy among students and teachers in three types of secondary schools),中西部社会学研讨会,明尼阿波利斯,MN。

这个研究探索老师和学生对"民主"意义的认识。这些老师和学生来自于一个新兴民主国家(捷克共和国)中三种不同类型的中学。该研究是在一个多维语义空间范围内对语义进行定位。1997 年的秋天,调查人员在捷克共和国进行了两个月的田野考察研究。在这两个月的时间里,完成了收集数据的工作。老师和学生代表分别来自一个大城市、两个中等城市和一个小镇上的三类不同学校。在他们完成一个有 57 对两极形容词组成的语义微分量表过程中,调查者要他们去考虑捷克共和国的政府类型。在研究中还采用了因素分析法,用一些形容词来表示一些少量的变量因素。结果表明,在评价性因素中,具有显著的年龄差异,在能力和稳定性因素中存在着性别差异。在评价、能力以及稳定性因素中,存在着学校级别的差异,而在稳定性因素中存在着社会等级/声望的差异,而在普遍性因素中则不存在任何显著的差异。

- Cogliser, Claudia C. , & Schriesheim, Chester A. (1994),一种测试语义微分两极化项目的新方法的设计和应用(Development and application of a new approach to testing the bipolarity of semantic differential items),《教育和心理测量期刊》(*Educational and Psychological Measurement*),54,pp. 594-605。

在该研究中,两极性这个新概念并不需要单维性,通过使用这个新概念,研究者开发了一种适合两极性的语义微分量表的测试方法。使用多维测量和 63 个大学生被试对"Fielder 最少首选合作者"工具进行评价,揭示出这个多维测试与两极性具有显著的分离性。

- Hayashi, Naoki, Yamashina, Mitsuru, Ishige, Naoko, Taguchi, Hisako, Igarashi, Yoshito, Hiraga, Masashi, & Inoue, Yukiyo. (2000)。精神分裂症患者与其临床医生的感知:评估治疗关系的语义微分技术的应用(Perceptions of schizophrenic patients and their therapists: Application of the semantic differential technique to evaluate the treatment relationship),《综合精神病学》(*Comprehensive Psychiatry*),41,pp. 197-205。

这项研究试图对治疗与精神分裂症病人的相互关系作出评估。这一关系可以通过考察病人和他们的医疗师对各自的知觉情况以及相互的知觉情况来得出。这些知觉被假设为反映了关系的特征。通过使用带有 17 对形容词的语义微分技术,对 158 名精神分裂症病人以及 11 名精神病医师的知觉进行了评估。每个医疗师都认为,他们作为治疗人员对病人有一种支持的关系。根据评估结果,研究者构建了 8 个具有高内部一致性的合成量表。知觉要素之间的相互关系可以由合成量表的得分之间的相

关性分析表现出来。这种相互关系似乎与研究者的临床经验有一致关系。对量表的因素分析研究产生了三个互不相关的因素,这些因素被认为具有治疗关系的特点。病人与治疗师之间的合作关系这一因素表明两个被试之间的信任程度,而治疗联盟的情感和关系方面也被假定如此。医疗师的消极性因素反映出医疗师的消极作用以及病人的临床治疗稳定性。病人的强度因素与病人的情景相关强度和性格强度有关。研究显示,病人和医疗师所做的评价对于目前的治疗关系评价也是有效和适用的。

关于语义微分法的推荐读物

Bishop, J. Joe. (1999, April). *Locating Czech democracy: A semantic differential analysis of the meaning of democracy among students and teachers in three types of secondary schools.* Paper presented at the Midwest Sociological Society Meeting, Minneapolis, MN.

Cogliser, Claudia C., & Schriesheim, Chester A. (1994). Development and application of a new approach to testing the bipolarity of semantic differential items. *Educational and Psychological Measurement*, 54, 594-605.

Hayashi, Naoki, Yamashina, Mitsuru, Ishige, Naoko, Taguchi, Hisako, Igarashi, Yoshito, Hiraga, Masashi, & Inoue, Yukiyo. (2000). Perceptions of schizophrenic patients and their therapists: Application of the semantic differential technique to evaluate the treatment relationship. *Comprehensive Psychiatry*, 41, 197-205.

Moss, Claude S. (2001). *Dreams, images and fantasy: A semantic differential casebook.* Ann Arbor, MI: Books on Demand.

Osgood, Charles E., & Luria, Zella. (1954). A blind analysis of a case of multiple personality using the semantic differential. *Journal of Abnormal and Social Psychology*, 49, 579-591.

Osgood, Charles E., Suci, George J., & Tannenbaum, Percy H. (1957). *The measurement of meaning.* Urbana: University of Illinois Press.

Schriesheim, Chester A., & others. (1994). The equal-interval nature of semantic differential scales: An empirical investigation using Fiedler's Least Preferred Coworker (LPC) scale and magnitude estimation and Case Ⅲ scaling procedures. *Educational and Psychological Measurement*, 54, 253-262.

Snider, James G., & Osgood, Charles E. (Eds.). (1969). *Semantic differential technique: A sourcebook.* Hawthorne, NY: Aldine.

6

GUIDE TO STATISTICAL ANALYSIS OF SOCIAL SCIENCE DATA

社会科学数据统计分析导读

《研究设计与社会测量导引》第6部分，有助于研究者寻找证明假设的统计工具的信息和利用个人电脑来收集、分析和解释数据。只有当研究者具备研究设计的必需知识，了解可以利用的所有统计工具和通过个人电脑使用这些工具的方法时，高级的研究才可能完成。实现了这些条件，最大效力的研究选择才开始激起创造的想像力。

定性变量和定量变量都需要恰当的统计工具来证明研究假设。本部分6.1的内容简明复习了统计检验，描述了参数统计与非参数统计之间的差异，以及如何利用它们从样本推论总体。同时，本节也简要地回顾了常用的组间检验与相关检验。

今天，社会科学中的计算逐渐开始使用高性能的个人电脑，它们的主机被用于大样本的数据分析和相关分析。在过去的几十年里，计算机技术得到迅速地发展，发明了高性能的台式和手提电脑。因为这种技术已非常普及，所以对研究者和研究生而言，精通计算机技术是必需的。随着方便获得信息的远距离互联网功能的出现，研究者们现在可以在户外工作，并且完成同样类型的分析工作，而先前这种工作需要搬到学校里去使用大型计算机。

学生们和研究者们期望使用这些工具，因此他们必须掌握基础的和高级的统计知识。计算机可以用于下列方面，例如，产生社会系统的正式模型；模拟个人、团体或者国家的行为；检索大量档案资料。

因为统计计划是构成研究整体所需要的部分，所以研究者越早关注这种研究活动越好。选择好问题、理论、研究设计和假设后，就应该开始做统计计划。研究者不能等到了收集或者分析数据过程中才做统计计划。这是因为，研究者如何分析资料，与所提出的研究问题及所收集的数据的类型是密切相关的。年轻而没有经验的研究者们，在任何实践工作或者田野工作开始之前，尤其应该做好统计计划。

6.1 质疑科学

下列文章应该会使研究者们对他们准备检验的假设的重要性的认识更加敏锐。正如,Lurie 陈述的:

> 正确地决定假设是什么,这些假设是关于什么的,确信所要达到的精确度是多少等等,都是科学家的责任……而且越来越多的科学家开始意识到他的责任,并且在工作中对它们加以考虑,因此,他的结论就更加正确有效,他解决的办法也更加接近现实。

傲慢的(impertinent)提问者:科学家到统计学家思维的指南

William Lurie

前 言

用统计的修饰来润色科学已经变得流行时尚。方程中至少有两个求和符号才是完备的,替代符 *ij* 依附于相似的 X_S、Y_S和 Z_S;"多形态分布"(polymodal distribution)、"逆相关"(inverse reciprocal correlation)和"多变项离差"(multivariate deviations)等惯用语可以在"如此这般的社会期刊"、"这样那样的协会集刊"和"等等等等的座谈会议议事录"这类杂志的任何一页上看到。

但是,除了为这些刊物提供数学的和语言的润色以外,如果统计学家是真正地帮助科学家,他必须完成一个既必需又恼人的任务:向科学家询问不适当的问题。的确,如果直率地询问问题,这可能显得不仅不适当而且是不正当地探问——因为他们涉及科学家思想的基础。通过这些问题,在基础中不被怀疑的弱点可能显露出来,而且揭露某人思想中的弱点是一个相当不愉快的事情。

如果统计学家不但精通统计学,而且精通人类交流的方式,那么,他就会艺术地询问这些问题,或者甚至一点也不把它们当作问题来问。他能很好地指导科学家讨论,不用明确地询问问题,来让问题的答案自动出现。

幸运的是,如果某人同时具有科学的和统计学的修养,那么,科学家的统计学良知就是向科学家提出这些问题,而不是不适当地质问是否存在科学探索精神。

那么,不论这些问题是出现在科学家自己的思想中,还是思想外,它们是什么呢?它们是:

- 对于你正在进行的实验,你的想法是什么?
- 这些想法所涉及的科学领域,是关于哪个方面的?

- 你对所设想的这些想法的准确性有多大把握?

为了理解统计学家提问这些问题的理由,首先,让我们看看统计学家是如何看待科学家的活动的。

从统计学家的观点看,科学家所做的就是:完成实验或者观察,以获得一个关于科学家感兴趣的那部分世界的结构的想法的数据,因此,科学家可以决定他的想法是否正确。

就科学家活动的这些楷体字方面的每一个部分,存在一个相对应的问题。

接着,让我们检查科学家活动的这些方面的每一个部分,其目的和关于它的问题的结果。

科学家的想法

傲慢的提问者必须冒着暗示科学家没有考虑清楚的危险。当然,即使产生这种结果的暗示并不是打算用来使暗示者受被暗示者情感的影响。但确切地,这个暗示,也许清楚地与"你的想法是什么"这个问题有关。

为什么统计学家会问这个傲慢的问题呢?因为这是统计学家能够帮助科学家完成他的目标的前提。一个朦胧公式化表示的想法不仅能够被讨论,最多是有困难地被讨论,而且更进一步说,要测试它的正确性实际上也不可能。因此,统计学家有个规则,称为:明确的假设。这个规则表明了这样一种要求:由实验决定其正确性的想法最好在实验进行之前就清楚、详细地陈述出来,并且尽可能采取明确的形式。这个想法可以和一个因素的影响相关,可以和一些因素的影响相关,也可以和正在实验的具有数字特征的属性相关。在一项研究的初期,所寻求的是有影响力的因素(如,那些当它们在变化的层次上,能引起足够的不同的结果),想法(或者假设)可以是不具体的,但必须是明确的。假设可以是广泛的,但它必须是明确的广泛,换句话说,即使它不是一个关于细节的假设,但它的边界必须清楚地描绘出。

例如,"因素,A、B、C 和 D 分别影响结果","因素 A 和 B 共同影响结果,且与所期望的因素 A 和因素 B 的单独影响不同","因素 A、B 和 C 共同影响等等"。或者研究末期,更加具体化,"因素 A 在 α_1 水平的测量结果为数字值 $N \pm n$"。

为了准确无误地强调明确假设的要求,让我们使用一个明显夸大的例子来处理特别的主题:一个工业心理学家的任务是去发现为什么会计办事员会出现太多附加的错误(决定多少错误"太多"是另一个统计问题,在这里将不予考虑)。

为这个例子的目的,心理学家可能对自己说:"作为一个心理学家的经验告诉我,一个人所处的情境会影响他的行为。因此,让我来找出引起办事员犯这些错误的情境是什么。"如果心理学家想法的形成不再进一步深入,那么,他显然会不停地去试图找出情境是什么,从现在直到永远,因为"情境"是没有边界的。

例如,情境可能不仅包括办事员的工作情况,还包括他们的家庭情况,他们的童年经历,他们的理想生活,而且情境的可能性是无限的。假设尚未足够明确地公式化表达出,而且其情境也没有被清晰充分地描述出,那么显然地,得出假设正确的决定是不可能的。

但现在,心理学家的统计意识觉醒了,而且他的想法也开始从它们最初的散漫中变得具体。“情境?——好吧,更详尽些,我们只考虑办公室情境,且在办公室情境中,我将选择三个我认为会影响办事员工作的因素。我选择研究的影响因素是:温度、湿度和噪音。我的明确假设是:就会计办事员犯附加错误的数量而言,不同水平的温度、湿度和噪音是有影响的。”这个假设已经能够(或者很可能)更加明确地表达出来(如,包括照明度、每个雇员的办公桌空间等因素),但是统计优点的路径方向已经指出,进一步沿着这条路径前进就留给读者。

现在,假定假设已经足够明确地用公式表示出来,科学家和统计学家就可以一起重新检查实验计划(或者设计),且他们确信,当能充分决定科学家想法正确(或者错误)时,数据也就获得了。

科学家感兴趣的那一部分世界

其次,在问“你的想法是什么”时,傲慢的提问者必须小心谨慎。尽管某人可能承认,他的想法不如他所希望的那样清楚而明确地用公式表示出来,但是,问题“你的想法是什么”毕竟带有这种倾向。对被提问者而言,他所不清楚的思想主题含意,肯定不是一个言过其实的含意。统计学家有含蓄批评科学家的自负的理由。统计学家对科学家的理由可以清楚陈述为:“这只对你自己有利。如果要我帮助你做决定,那么,根据实验事实,不论你的想法是否正确,我都必须尽可能明确知道你的想法是什么,且这些想法是关于什么的。我称这个必要条件为:模型的公式化表达”。严格地说来,模型的公式化表达确定了必要条件,即科学家的想法是只适用于实验条件(小范围的应用),还是适用于除了实验进行下的那些具体条件(大范围的应用)之外的条件(如,因素的不同层面),这两者之间存在明显的差别。为什么这种差别是必要的呢?因为,当获得实验数据时,根据假设是希望适用于大范围还是小范围,可以以不同的方法进行数据分析。

让我们回到工业心理学家的例子。他的实验条件是:温度,40 ℉、55 ℉和 70 ℉;湿度,40%、50%和 70%;噪音,40 分贝、50 分贝和 70 分贝。

通过分析实验数据获得结论这一方法可能很好地区分科学家和统计学家(例如,关于想法正确的结论),且不论能否获得任何结论,如果可以,那么它们是什么,取决于科学家是想把他的结论只适用于已经用于实验的三个标准的温度、湿度和噪音,还是也想适用于其他(不具体的)温度、湿度和噪音。支持小范围结论的数据可能不够支持大范围结论。因此,科学家必须清楚他的假设是关于什么的,且必然地也清楚他的结论是宽的还是窄的。而且,对科学家而言,统计学家努力确保科学家的确清楚这一点,大概会显得不恰当。

科学家的想法正确与否

统计学家的第三个问题——“你对所设想的这些想法的准确性有多大把握?”是三个问题中最不重要的。不像其他两个问题,这个问题不去探查科学家思想的基础,但要求科学家去量化预先没有量化的方面(事实上,这个要求与科学家自己对定量数

据的偏好是相一致的)。这个方面涉及确信水平,通常语言给我们提供的是定性描述的术语(略微确信,基本确信,相当确信,非常确信)。但是这些术语对科学的使用而言是不够明确的。因此,统计学家向科学家请教以决定和表达他所期望的确信水平的定量术语,这样通过定量数据分析,不论所期望的结论的确信水平是否实现,确信水平都可以被确定。统计学家把期望的确信水平的选择和定量表达称为显著性水平选择(significance level selection)。那么统计学家如何帮助科学家选择期望的确信水平呢?通过回到科学家意识的最前面,即他没有察觉到的事件固有变化性的意识(如,因为独立变化,不重复的实验将精确地给出相同的结论);通过帮助科学家决定尚未被"证实"、仅是独立变化实施过程的假设所期望的确信;根据实验数据,通过提供数学的工具以决定期望的确信水平能否实现。譬如,在温度—湿度—噪音的实验中,当所有的数据都收集起来,且科学家准备分析它们时,他会决定他的假设正确与否,统计学家将告诉科学家:"当然,你知道的,假使你在尽可能接近的条件下完成实验,你会得到稍微,甚至有些不同的结论。这个结论可能(或仅仅偶然)使你相信温度的确影响结论的正确,即使它事实上不是;或者即使你没有完成实验,只是恰好完成那个特别的实验,可能也就是它的数据使你相信温度存在影响,即使它事实上不是。但是,我可以检验你的这些数据。我可以确保,当你陈述结论,温度的确影响正确性时,你将会有**5%**,或者**1%**,或者**1‰**错误的概率,因为我会告诉你这是不太可能的。那么,你想采用哪种概率?当事实上不是时,你选择了一个非常小的犯错误的概率来说明存在温度影响,但可能事实上存在时,你选择了大的犯错误的概率来说明不存在温度影响。我也可以为你计算出这个概率。那么,你想采用哪种概率?"

当科学家已经在他的结论里选出他愿意采用的犯错误(或者,相当于,他想他的结论是正确的把握性有多大)的概率时,统计学家可以分析数据,并且告诉科学家他可以正当地得出什么结论(如,他可以得出关于他的想法是正确的决定)。

结 尾

总之,统计学家的责任是去询问,而不是回答这些问题。科学家的责任则是确切地决定假设是什么,这些假设是关于什么的,对它们的准确度有多大把握。

统计学家通过提出鲁莽的问题,正好明确地引起科学家注意自己的责任,这些责任可能是科学家没有意识到的。当越多的科学家开始意识到他的责任,并在他的工作中考虑这些责任,那么他的结论就越准确、有效,且他的处理也越接近现实。

资料来源:William Lurie (1958), "The Impertinent Questioner: The Scientist's Guide to the Statistician's Mind", *American Scientist*, 46, 57-61, 使用已获得授权。

6.2 测量的四种层次及其相对应的统计

本书的第7部分包括测量各种社会和行为的变量的量表。这些量表以定类、定序、定距和定比类型的数据为基础。测量的这种分类是由 S. S. Stevens(1951)第一次提出的。

表6.1 测量的四种层次

层次	特　质	例　子	定义特征
定类	以标签分类	喜好(喜欢或讨厌) 投票记录(同意或反对)	每个观察值都属于它自己的类别
定序	某一连续体中的相对值	大学中的等级 相对于他人的社会等级	一观察值高于或低于另一个
定距	点间相同距离	智商 温度	在同一间距的测量中某一个值不同于另一个
定比	绝对零点	年龄 体重 时间	定比值具有意义,因此一个数值可以说是另一个的两倍或其他比率,而且在数量测量中没有的存在是可能的

- 定类或类别量表(nominal or classification scales)使用标签、数字或其他象征物仅仅是为了对物体、人或特征(如,高或低,农村或城市)进行分类。这种测量层次的特征是,其变量在本质上属哲学范畴。例如,社会阶级(按分类)、性别和邻近地区都可以被认为是定类测量层次的变量。
- 定序或等级量表(ordinal or ranking scales)包括一种测量层次,它包含在某种与类别相关的事物的各种分类的排序中,例如,阶级等级高或低。设想一群相同阶级的群体,如果其中某些阶级小群体之间关系比一些但不是所有阶级强,我们就部分地排列了量表;如果关系比所有的阶级都强,那么来进行完全的阶级排列,这就是定序层次。
- 定距量表(interval scales)具有定序量表的所有特征,此外,它还要求量表上任何两个数字之间的距离是已知的(例如,温度计的摄氏度)。较之定序层次,定距层次的测量更适合用定距量表。Thurstone 的同等表现定距量表就是一个例子。这种测量层次具有许多社会的和行为的评价特征。有时,如何分派测量也是有一些自由的。
- 定比量表(ratio scales)具有定距量表的所有特征,此外,它们具有真正零点(例如,考试中的分数)。任何两个点之间的比率是独立于测量单位的。定比量表很少用于社会或行为科学,因为人或机构不会存在真正零点的特征。

每个测量量表都有定义关系,这使得其更合适于特定统计检验。一般地,定类层

次和定序层次要求用非参数检验,然而定距层次和定比层次却要求用参数检验(也有例外,比如定距数据使用非参数统计)。因为大部分的指标和量表都是定序的,因此,非参数检验显得特别重要,而且给量表的定义特征配以适合的统计也是很必要的。

关于测量层次和研究设计中变量的选择,我们有必要牢记两点:第一,随着量表类型从定类到定比,测量的精确性也递增。假如定类测量层次可能允许我们得出两个组是不同的话,那么定比测量层次则允许我们得出不同的程度。第二,研究者断言一个变量在定距测量层次可能是不科学的,但人们对于这种断言倾向于保持稍微自由的态度。

统计显著性的意义及其用法 6.3

不论你是用参数方法还是非参数方法,推论统计(相对于描述统计)的目标都是为了得出这样的结论,即结果的概率是归因于偶然性而不是一些假设的原因。

在这个基本原理上,统计显著性是一个中心主题。下面的摘录来自 Thompson,讨论这个概念及应用的一些最重要方面。

关于统计方法的推荐读物

Bechhofer, Robert E., Santner, Thomas J., & Goldsman, David M. (1995). *Design and analysis of experiments for statistical selection, screening, and multiple comparison.* New York: John Wiley & Sons.

统计显著性检验的概念

Bruce Thompson

很少研究者知道统计显著性检验做什么,不做什么,所以他们得出误解的结论。甚至更通常地,研究者知道统计显著性检验的要素,但是无法把它融进他们的研究。例如,研究者可能知道样本规模对统计显著性的影响,但是,当用于解释一个有着几千个对象的研究结论时,这种洞察力就没有被表达出来。

这篇文章将帮助你很好地理解显著性检验的概念,它讨论了概率的意义,显著性检验的概念,反对显著性检验的观点,误解或者替代物等等。

统计显著性检验中的概率是什么

无论是来自总体的一个随机样本,还是近似于随机的代表性样本,研究者都可以用统计显著性检验。在确定先决可接受的由抽样误差引起的推论误差的概率(从0

到 1.0)时,统计显著性检验要求主观判断,抽样误差是从很多不同的抽样过程中固有的“偶然”中获得样本时产生的。只有从完整的总体中获得数据,抽样误差才会消除。

一个概率(p),当虚无假设(null hypothesis)在总体中确实真实时,决定拒绝这个虚无假设的概率(如,一个假设为均值$_1$ = 均值$_2$ = 均值$_3$,或者 $R^2=0$),称作“α”,也叫 $p_{(临界的)}$。当我们选择一个 α 水平时,做出错误决定的最大概率,称甲种误差(type I error)。因为,一般 α 是很小的,所以,这种误差的概率也将很低。于是,$p_{(临界的)}$是以主观判断为基础选出的,这种主观判断是关于在一个特定的研究条件下甲种误差的结果将是怎样,以及关于这些结果特定的个人价值观的。

第二种概率,$p_{(计算的)}$(如所有的 p's 一样,从 0 到 1.0)是计算出的。概率只有在假设足以抑制计算以致于一个已知的问题只有一个答案的情况下才可以计算出。

在两个样本中,获得 IQ 均值是 99 和 100 的概率是怎样的? 首先,这取决于从样本中选出的总体的实际统计参数(如,均值)。如果总体均值分别是 99 和 101 时,这两个样本统计值(均值$_1$ =99;均值$_2$ =101)将是最可能的(放弃最大的 $p_{(计算的)}$);如果总体均值都是 100,这两个样本统计值将是最不可能的(放弃最小的 $p_{(计算的)}$)。由于实际总体参数值是未知的,那么我们必须假定参数值是什么,在统计显著性检验中,通过虚无假设来假定参数值是正确的。换言之,我们为这些计算假定虚无假设恰好是真实的。

影响 p 值计算的第二个因素包括样本量。样本量越小,较总体(偶然的)而言,样本(统计值计算的)的代表性潜在的也越小。举个例子,选出样本量都是 5 的两个样本,比样本量都是 50 的两个样本更可能放弃“偶然的”统计值(均值,r's 等。)因此,$p_{(计算的)}$的计算也必须考虑样本量的影响。如果样本量都是 5 的两个样本的均值是 100 和 90,并且样本量都是 50 的两个样本的均值也是 100 和 90,那么,均值是相等的虚无假设将为大样本放弃最小的 $p_{(计算的)}$值。因为,假定虚无假设完全是真实的,随着样本量的增加,不相等的样本统计值的概率是减少。总之,$p_{(计算的)}$值的概率提出这样问题:

- 假定一个样本数据来自一个总体,它的虚无假设恰好是真实的,在给定的样本量下,某人获得自己样本数据的样本统计值的概率是多少?

即使没有计算这个 p 值,我们也可以对 $p_{(计算的)}$做出合理的判断。在下列成对的研究中哪一对的哪一个的 $p_{(计算的)}$值会较小?

- 两个研究,每个都包含三组 30 个样本:第一个研究的均值是 100,100 和 90;另一个研究的均值是 100,100 和 100。
- 两个研究,每个所包含的两组样本因变量的标准差都是:$SD_1=4$ 和 $SD_2=3$,但是第一个研究的样本量是 100 和 100,而第二个研究的样本量是 50 和 50。
- 两个研究,每个都包含一个多元回归,用 X_1、X_2和 X_3预测 Y,且样本量都是 75,第一个研究的 $R^2=0.49$,而第二个研究的 $R^2=0.25$。

统计显著性实际上要告诉我们什么

统计显著性提出下列问题:

> 假定一个样本数据来自一个总体,它的虚无假设正好是真实的,并且提供我们样本统计值和样本量,那么我们抽样结果所计算出的概率会比甲种误差可接受的极限 $p_{(\text{计算的})}$ 少吗?

当 $p_{(\text{计算的})}$ 小于 $p_{(\text{临界的})}$ 时,我们使用一个决定规则,即我们将"拒绝"虚无假设。拒绝虚无假设的决定被称为"统计上显著"结论。根据我们的假定,包括虚无假设完全是真实的假定,所有决定方法是让我们相信我们抽样结果相对地不太可能。

然而,虽然得出 $p_{(\text{临界的})}$ 是容易的,但计算 $p_{(\text{计算的})}$ 却是困难的。照惯例,统计值检验(如,F,t,X^2)已经被等同于 p's 的重复表达。因为统计值检验$_{(\text{计算的})}$(简称 TS)是比较容易获得的。完全等同于已知的 $p_{(\text{临界的})}$ 的 $TS_{(\text{临界的})}$ 可以从广泛可使用的表格中获得,α 和样本规模可以在制成表格的数值中找到。不同的 $TS_{(\text{计算的})}$ 是依据于被检验的假设而计算出来的。我们所决定的统计值检验规则惟一不同的是:当 $TS_{(\text{计算的})}$ 大于 $TS_{(\text{临界的})}$ 时,我们拒绝虚无假设(称作"统计上显著"),然而,比较一个给定数据的 p's 和 TS's 经常会得出同样的结论。

记住,了解抽样结果是不太可能的,假定虚无假设是真实的,这可能是没有用的。一个不大可能的结果不是一个必要的重要的结果,正如 Shaver(1985,p58)假设的两个教师之间的对话:

> Chris:……如我的导师所建议的,我确定显著度是 0.05。如果群体不是真的不同,那么存在一个差异,即大样本将少于 5% 偶然地出现。像这样一个不太可能的发生应当是重要的。
>
> Jean:等一下,记得那天你到办公室打电话回家吗?正当你拨完电话号码时,你的小孩拿起电话打电话给某人。因此,电话没有接通,你就接通了另一个人的电话……那么,这想必真是一个重要的事件啦?

为什么不使用统计显著性检验

统计显著性检验需要缺少相对应利益的努力的投入。科学是一项发生在指定条件下的孤立关系的工作,因此,知识是被创造出来的,且是可以累积的。但是统计显著性没有充分地提出——一个指定研究的结果是否将重复(Carver,1978)。作为科学家,我们必须问两个问题:第一,抽样规模的影响是什么;第二,这些结果是否将被泛论。统计显著性检验没有回答任何一个问题(Thompson,出版中),因此,统计显著性检验可能从更重要的考虑中转移注意力。

统计显著性检验的误解

在统计显著性检验现代使用中出现的许多问题始于研究者使用的语言。数个名称可以表达一个单独的概念(如,"$SOS_{(\text{中间})}$" = "$SOS_{(\text{解释})}$" = "$SOS_{(\text{模型})}$" =

"$SOS_{(回归)}$"),且术语在不同情境中有不同的意思(如,"单一变量"意味着只有一个因变量多个潜在自变量,也可以指一个只由一个单一变量计算而得的统计值)。

克服语言上的三种习惯将有助于避免没有意识到的误解:

- "统计上显著"(statistically significant)而不是"显著"(significant)。将这个概念称为一个短语将有助于分清拒绝一个虚无假设和获得一个重要结论之间的不正确的关系。
- 请不要说"我的结论达到统计显著"(my results approached statistical significance)。这个说法在统计显著性检验逻辑的前后关系中是毫无道理的。对这个问题,我最喜欢的回答是由编辑们提供的,"你是如何知道你的结论没有设法避免统计上显著?"(how did you know your results were not trying to avoid being statistically significant?)
- 请不要说"统计显著性检验评判结论是否'出于偶然'"(the statistical significance testing evaluated whether the results were"due to chance")。这个说法给出了一个印象:可重复性是由统计显著性检验来评判的。

哪种分析更适合统计显著性检验

应该强调统计显著性检验的两种分析(*Journal of Experimental Education*,1993)。首先,在所有分析中都应该计算和解释效果大小。这些可以是 r^2 类型的效果大小(如,R^2,eta^2,Ω^2),它们评判分析中所解释的方差比例,或者统计中标准差(如,均值的标准差),或者两者兼有。

其次,结论的可重复性必须经验地调查,或者通过研究的现实重复,或者通过使用诸如交互确认(cross-validation),折叠法(jackknife),引导程序法(bootstrap)等方法(参见 Thompson,出版中)检查。

资料来源:Bruce Thompson(1994),"The Concept of Statistical Significance Testing," *Practical Assessment, Research & Evaluation*,4(5). Retrieved September 3,2001,使用获得授权。

6.4 总结相关与群体差异的共同测量法

统计方法使我们能够研究且精确地描述平均值、差异和关系及决定结果是趋于变化或者一些其他因素(例如,自变量的操作化)。

在过去的30年里,可利用的统计检验方法的数量多了很多,以致就是专业的统计学家也不能掌握所有的变化(Kanji,1993)。由于这些检验已经变得越来越多,精确度也提高了,可以检验的假设种类的数量也增加了。

在第6部分的后面两节,我们将复习用于检查变量间相关与群体间差异的一些共同的检验法。接着,我们将转移到最一般和最简单的计算说明:Pearson 相关(r_p)、t 值和 F 检验。计算和使用这些统计值的介绍论述可以是手算的,也可以是计算机算的。

在所有这些情况里，希望能很好使用和理解这些统计工具的学生将需要在这些领域里接受一定的课程训练[关于这些检验的进一步信息，请参见 Salkind(2000)]。

参考文献

Kanji, Gopal K. (1993). 100 *Statistical tests*. Thousand Oaks, CA: Sage.

Salkind, Neil. J. (2000). *Statistics for people who (think they) hate statistics*. Thousand Oaks, CA: Sage.

相关测量法概述 6.5

一般的相关测量法包括 Pearson 积距相关 r，相关比率 eta，gamma(Γ)，Spearman 等级差异系数 rho(ρ)，lambda(Λ)和复相关 R。

表 6.2 总结了常见的相关测量法。因为科学调查的主要目标是揭露关系，所以对科学家而言，学习如何使用这些测量法就成为标准的训练。

表 6.2 常见的相关测量法

相关测量法	符号	测量层次	何时使用
Pearson 积距相关，r	r_{xy}	定距	测量连续的、线性关系的两个变量之间的关系。当以大量观察结果为基础时，相关关系是最可信的
相关比率，eta	η_{xy}	定距	测量连续的、曲线关系的两个变量之间的关系
Spearman 等级差异系数，rho	ρ_{xy}	定序	测量两个等级之间的关系。主要用于个案或两个变量等级排列，那么每个变量的等级从 1 到 N 排列
Lambda	λ_{xy}	定类	测量双变量分布的、定类的两个变量之间的关系
复相关系数，R	R_{123}	定距	测量由若干连续的自变量和因变量形成的组合之间的可获得的最大关系
偏相关系数，r	$r_{12\cdot3}$	定距	测量结果由第三个连续变量固定引起的两个连续变量之间的关系
双连续相关系数	r_{bi}	定序	测量一个二分变量和另一个连续变量之间的关系，且假定二分变量是连续的、常态的
点双连续相关系数	r_{pbi}	定类	测量二分变量和连续变量之间的关系
列联系数	C	定类	测量可以被分为两个或多个种类的，且它们不是定量的两个变量之间的关系
ϕ 系数	ϕ	定类	测量都是二分变量的两个变量之间的关系

6.5.1 Pearson 积距相关系数

积距相关系数是通过检验由变量贡献的方差比例而计算出的。原始值公式是:

$$r_{xy}=\frac{N\sum XY-\sum X\sum Y}{\sqrt{[N\sum X^2-(\sum X)^2][N\sum Y^2-(\sum Y)^2]}}$$

r_{xy} = X 与 Y 之间的相关系数

N = 样本大小

X = 变量 X 的个别值

Y = 变量 Y 的个别值

XY = 每个 X 值与所对应的 Y 值的积距值

X^2 = X 个别值的平方

Y^2 = Y 个别值的平方

(1)r_{xy}的解释

任何类型的相关都容易被误解,因为相关经常被看作等同于因果关系。两个变量可以同时发生变化的事实经常被误解为:一个变量的变化引起了另一个变量的变化。这是初学研究者经常犯的一个严重错误。以下是解释相关系数的一些别的指导原则。

0.00 - 0.20	很少或者没有关系
0.20 - 0.40	少量关系(弱关系)
0.40 - 0.60	大量关系
0.60 - 0.80	很有用的关系
0.80 - 1.00	强关系

一个相关的意义是由它的大小决定,而不是它的符号。符号与关系强度无关,它只决定关系的方向:正向或直接的,负向或间接的。

相关系数的平方,就如决定系数(r_{xy}^2),提供了意义或者相关系数重要性的最好提示。它确认由另一个变量的方差说明一个变量方差的数量。一个系数为 1.0,或者拥有完全相关的积距相关 r_{xy} 的决定系数是 1,或者 100%,且一个拥有一半强度相关(r_{xy} = 0.50)的积距相关的决定系数是 25%。一个重要的相关意味着数据所来源的总体中的变量很可能不是零相关:即,它们是相关的。

(2)使用计算机来计算 Pearson 积距相关

对行为科学和社会科学而言,SPSS 是最重要的统计和数据分析的软件包之一。本书中所设计的使用计算机来计算统计值的例子将都是以 SPSS 为基础的。图 6.1 中所举的例子使用的是 20 个观察结果的简单数据,两个变量:收入(美元)和教育水平。双变量相关的结果由以下的数字表示。

相 关

		收入	教育
收入	pearson 相关	1.000	0.574**
	双尾检验的显著性概率	.	0.008
	样本量(N)	20	20
教育	pearson 相关	0.574**	1.000
	双尾检验的显著性概率	0.008	.
	样本量(N)	20	20

** 相关是在 0.01 水平显著的

图 6.1 相关系数计算的 SPSS 输出结果

均值差异检验的概述 6.6

组间差异检验的常见测量法，包括独立或者配对均值的简单 t 检验和评价两个均值以上差异的 F 检验。F 检验(以 R. A. Fisher 命名)也称作方差分析(ANOVA)。

6.6.1 独立均值检验的 t 值

t 值是通过比较组间方差的贡献和组内方差的贡献计算而得的。两个不同样本都被检验的独立样本检验的原始值公式如下：

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \times \frac{n_1 + n_2}{n_1 n_2}}}$$

$\overline{X}_1$ = 样本 1 的均值

$\overline{X}_2$ = 样本 2 的均值

n_1 = 样本 1 的个案总数

n_2 = 样本 2 的个案总数

s_1^2 = 样本 1 的方差

s_2^2 = 样本 2 的方差

6.6.2 配对均值检验的 t 值

配对均值的 t 值是通过比较主题间方差的贡献和主题内方差的贡献计算而得的。一个样本检验两次的配对样本检验的原始值公式如下：

$$t = \frac{\sum D}{\sqrt{\frac{n\sum D^2 - (\sum D^2)}{(n-1)}}}$$

$\sum D$ = 样本间所有差异的总数

$\sum D^2$ = 样本间差异平方的总数

n =样本量

(1)t 值解释

在所有情况下,一般地,t 值是一个检验值,它一般与一个规定表的值或者一个期望值比较,假定下面的这个虚无假设为真:

$H_0:\mu_1=\mu_2$ (独立均值情况)或者

$H_0:D=0$ (配对均值情况)

当所得的值超过否定值或者规定表的值,可以得出所得的值独立出现的概率是非常小的结论。另一方法是 t 值被解释为使用诸如 SPSS 统计软件包的结果,决定所得值精确相关的概率。

(2)使用计算机计算 t 值

就图 6.2 所举的例子而言,使用一个简单的数据,由 60 个观察结果(来自两组独立的样本)和因变量——态度组成。图 6.2 表明了独立均值比较的 SPSS 输出结果。

图 6.2 表明了 SPSS 输出的概率。在这个例子中,这两个样本差值预计变化的准确概率是 0.891。这种结果是可以说明:以前使用显著度和否定值的惯用规定表的值来求得概率,但有统计软件后就不需要了。

6.6.3 两个均值以上的 F 检验计算

计算两个均值以上组间或组内差异的全部检验的 F 值,是通过比较贡献于组间方差和贡献于组内方差而得的。F 检验的原始公式如下:

$$F = MS_{组间}/MS_{组内}$$

$MS_{组间}$ = 由组间差异引起的变化

$MS_{组内}$ = 由组内差异引起的变化

(1)F 值的解释

一般地,t 值是与一个规定表的值或者一个虚无假设是真实的期望值比较而得到的,或者是检验值。这个虚无假设是:H_0: $\mu_1=\mu_2=\mu_3$。当所得的值超过否定值或者规定表的值,可以得出所得的值独立出现的概率是非常小的结论。另一方法是 F 值被解释为使用诸如 SPSS 统计软件包的结果,决定所得值的精确概率。

Independent Samples Test										
		Levene's Test for Equality of Variances		t test for Equality of Means						
		F	*Sig.*	*t*	*df*	*Sig.* (2-tailed)	*Mean Difference*	*Std. Error Difference*	95% Confidence Interval of the Difference: *Lower*	95% Confidence Interval of the Difference: *Upper*
ATTITUDE	Equal variances assumed	4.994	.029	-.137	58	.891	-1.00E-01	.73	-1.56	1.36
	Equal variances not assumed			-.137	47.635	.892	-1.00E-01	.73	-1.57	1.37

独立样本检验										
		方差齐次性检验		均值齐次性 *t* 检验						
		F	*Sig.*	*t*	自由度	双尾检验的显著性概率	均差	差值的标准误差	差值的95%置信区间：下限	差值的95%置信区间：上限
态度	假设方差相等	4.994	0.029	-0.137	58	0.891	-1.00E-01	0.73	-1.56	1.36
	假设方差不相等			-0.137	47.635	0.892	-1.00E-01	0.73	-1.57	1.37

图 6.2　独立均值 *t* 检验的 SPSS 输出结果

(2)使用计算机计算 F 值

就这一节的例子而言,使用一个简单的数据,即60个观察结果(来自三个独立组样本)组成,其分组变量是节目时间,因变量是青少年犯罪减少。图6.3表明了三个独立均值的全体比较的SPSS输出结果。

单因方差分析(ANOVA)

犯 罪

	平方和	自由度	均 方	F 值	显著性概率
组 间	1 133.067	2	566.533	8.799	0.001
组 内	1 738.400	27	64.385		
总 计	2 871.467	29			

图6.3 单因方差(ANOVA)的SPSS输出结果

6.7 因果关系和多变量分析:从单变量和双变量问题到社会行为的多变量分析

普遍认为,一个结果只存在一个原因,且假如发现若干个原因,那就假定一定存在不止一个结果。这种社会理论存在一系列的主要陈述:主张社会变化的惟一原因只有一个因素。这些观念被称作"决定论",它们包括地理学的、物理学的、种族学的、心理学的、宗教的、政治的、经济的、科技的和家庭的因素等等。

决定论所有这些观念的特征是主张惟一因素依据其内在的规律进行,与所有因素,包括人类意愿和愿望无关。这些单一因素理论相对地容易理解和同样吸引学者和门外汉。他们似乎从由社会问题引起的"负责"现象中获得真理,但过于简单化,单一因素理论歪曲了现实,例如,种族决定论给世界每个国家带来了偏见和歧视。

虽然,单一因素理论仍然盛行,但现代人更有判断力。现代方法包括允许和期望一个结果有许多不同原因(一果多因)。

6.7.1 原因的四种特征

原因可以出现连续、集中点或者聚集、产生分散结果和复杂网络等特点。

- 原因以连续物出现,就像链子上的环。这些原因有一些是直接的、即时的,其他则是间接的、遥远的。因此,以高度重复的方式在大公司里完成许多工作的直接事实,可能导致工人的动机和个人责任感的下降。间接原因是工厂制度和大众市场,这反过来是由蒸汽机、电动发动机和机械工具引起的。
- 若干原因可能集中产生一个变化。因此,电力交通工具和通讯发明集中增加了工业的分散。这些集中的原因经常被称为一个聚类(cluster)。
- 单一原因的结果可能会分散到社会的许多不同部门。因此,美国人获得的正式教育的平均增长对家庭、教堂、社区、军事组织和劳资关系有许多不同的影响。

- 集中和分散的现象可能与随后产生的原因的复杂网络的现象有密切关联。这是个非常普遍的显然特征,但是复杂性可以由原因在重要性上不同的认知来简单化,而且,重要的原因可能通过大部分观察到的结果的说明来确认。

6.7.2 多变量分析的统计世界

如今,多变量分析已经发展出能够同时处理三个以上变量或者属性的程度。使用这一类型分析试图把现实生活情境中的复杂变量解释清楚,这取决于那些在仔细研究下可以很好显现出本质关系的东西。多变量分析可以为预想问题提供增加的精确度(许多预测变量和衡量标准的关系),对干扰或者混淆变量提供更强的控制(保持更多变量不变),以指导原则指导态度量表、评价量表、心理测验和量表测量法的发展(发现行为的维度)。一些重要得多变量分析方法包括如下:

- 复相关和聚类分析(multiple correlation and classification analysis)
- 路径分析(path analysis)
- 因子分析(factor analysis)
- 偏相关分析(partial correlation analysis)
- 方差和协方差分析(analysis of variance and covariance)
- 多因判别分析(multiple discriminant analysis)

对这些方法的完整介绍不是本书所关注的范围和目的。

复相关和聚类分析(6.7.3)、路径分析(6.7.4)和因子分析(6.7.5)的叙述,都是对社会科学研究中很流行的多变量分析形式提供的介绍。

关于多元变量分析的推荐读物

Algina, James, & Keselman, H. J. (1997). Detecting repeated measures effects with univariate and multivariate statistics. *Psychological Methods*, 2, 208-218.

Bray, James H., Maxwell, Scott E., & Cole, David. (1995). Multivariate statistics for family psychology research. *Journal of Family Psychology*, 9, 144-150.

Grimm, Laurence G., & Yarnold, Paul R. (Eds.). (1995). *Reading and understanding multivariate statistics.* Washington, DC: American Psychological Association.

Johnson, Richard Arnold, & Wichern, Dean W. (1998). *Applied multivariate statistical analysis* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Mulaik, Stanley A. (1993). Objectivity and multivariate statistics. *Multivariate Behavioral Research*, 28, 171-203.

Sheu, Ching Fan, & O'Curry, Suzanne. (1996). Implementation of nonparametric multivariate statistics with S. *Behavior Research Methods, Instruments and Computers*, 28, 315-318.

Stevens, James. (2001). *Applied multivariate statistics for the social sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Stevenson, Jim. (1993). Multivariate statistics: V. The use of factor scores in psychiatric research. *Nordic Journal of Psychiatry*, 47, 169-178.

Streinger, David L. (1993). An introduction to multivariate statistics. *Canadian Journal of Psychiatry*, 38, 9-13.

6.7.3 复相关、回归和聚类分析

复相关($R_{1.234}$)是单个因变量的真实得分和从自变量的任何线性组合中获得的得分之间的一种简单相关。复相关类似于简单积距相关(r),它的统计值是从 -1 到 +1。系数(绝对值)越小,相关越弱;系数越大,相关越强。复相关可以通过它的平方来解释。R^2称作复确定系数,且表达了由回归方程解释的因变量的方差比例。

关于复相关、回归和聚类分析的推荐读物

Allen, Michael Patrick. (1997). *Understanding regression analysis.* New York: Plenum.

Bartolini, Stefano. (1993). On time and comparative research. *Journal of Theoretical Politics*, 5, 131-167.

Cortina, Jose M. (1993). Interaction, nonlinearity, and multicollinearity: Implications for multiple regression. *Journal of Management*, 19, 915-922.

Ganzach, Yoav. (1998). Nonlinearity, multicollinearity and the probability of Type Ⅱ error in detecting interaction. *Journal of Management*, 24, 615-622.

Morrow-Howell, Nancy. (1994). The M word: Multicollinearity in multiple regression. *Social Work Research*, 18, 247-251.

(1)应用范围

复相关 R 的使用已经为人所了解,但是,当多于 4 个或 5 个自变量(预报因子,predictor)被引入时,R 系数的计算就非常麻烦。计算机就可以消除这种局限,但是另一个局限就出现了,即只有当变量是连续的时候,系数才有用。多因聚类分析这一现代方法就可以消除这种局限。计算方法可以处理与定类(名义)测量一样的预报因子,且可以分析预报因子中任何形式间的,或者两个因子间或者一个因子与一个因变量间的关系。许多最有趣的分析问题包括若干因子变量的同时考虑(如,自变量)和它们对一个因变量的关系。有时,想知道所有变量加起来解释因变量变化的能力是多少,另一些时候,则分别看待每个因子如何与因变量相关,或者考虑或者忽略其他因子的影响,这是必要的。经常用于判断一个变量预测能力的标准是它对未解释方差或"误差"减小的贡献,另一种方法是看它的组均值不同于全部均值的程度。

一个不同但相关的问题是因子关系。人们有时会询问你想预测的标准是什么(如,特别的值或分数),或者其他拥有特定特征组合单元,而不是询问你能预测得多好。这是多元回归经常遇到的典型问题。最后,人们有时候想知道自己的预测能力是否比机会高得多。

多因聚类分析由 Frank M. Andrews, James N. Morgan, John A. Sonquist 设计发明的,而 Laura Klem(这节提到的)使用一种多变量方法,这与上面所有问题都有关联,且可能适合于许多种数据,传统方法的简单形式将不适合这些数据。相对于传统变量回归,多因聚类分析的最主要优点是更方便资料输入和集中于一组因子(例如,职业群)及这组因子间的相互关系的更易于理解的输出结果。

关于多元变量分析的推荐读物

Bartholomew, D. J. (1997). Fifty years of multivariate analysis. *British Journal of Mathematical and Statistical Psychology*, 50, 205-214.

Carroll, J. Douglas, Green, Paul E., & Anil, Chaturvedi. (1997). *Mathematical tools for applied multivariate analysis* (Rev. ed.). San Diego: Academic Press.

Hromco, Joseph George. (1997). A multivariate analysis of patient presenting problem: Why patients seek psychotherapy. *Dissertation Abstracts International*, 57, 7227B.

Ruth, Edward Keith. (1996). A multivariate analysis of gender, age, and ethnicity as elements shaping technological attitudes, anxiety, and self-efficacy in workplace settings. *Dissertation Abstracts International*, 57, 2449A.

Seilheimer, Thomas A. (1995). Self-efficacy, quality of life domains, and residential satisfaction: A multivariate analysis of mental patients. *Disseration Abstracts International*, 56, 1121B.

(2)虚拟变量回归分析

虚拟变量(dummy variable)都是二分变量的应用,当研究者遇到定类或名义变量时。计算机的2进制本质使它很适合于虚拟变量的计算,这表明,对每个单独回答出现(1分)或者缺失(0分)某种特征。对定性变量而言,这特别有用。举个例子,婚姻状况可以使用4个虚拟变量来介入:

已婚:1=是,0=否
未婚:1=是,0=否
丧偶:1=是,0=否
分居:1=是,0=否
离婚:0,0,0,0 模型将表明这个人是离婚的,不需要虚拟变量

如果某人是丧偶,他或她将有4位的值,而不是一个3分,他的4位虚拟变量的值是:0010。

所有这些虚拟变量将被归入一种典型的多元回归分析。当某个虚拟变量的某个值出现0时,即回归权数出现在回归方程中。否则,回归权数变成由研究者分配给特定状态的值。

虚拟变量可以以标准值或者原始值的形式作为自变量使用,且它们也可以被用于路径分析。如果一个研究者希望通过给变量值的特定组合"赋分"来发现在一个多元回归方程中期望的统计关系,那么,虚拟变量是有用的,否则,模型将假定变量间存在附加的关系。这些努力都称作虚拟变量回归分析。

关于虚拟变量分析的推荐读物

Eisinga, Rob, Scheepers, Peer, & Van Snippenburg, Leo. (1991). The standardized effect of a compound of dummy variables or polynomial terms. *Quality and Quantity*, 25, 103-114.

Figart, Deborah M. (1997). Gender as more than a dummy variable: Feminist approaches to discrimination. *Review of Social Economy*, 55, 1-32.

Hardy, Melissa A. (1993). *Regression with dummy variables.* Newbury Park, CA: Sage.

Herrera-Hernandez, Maria Lucila, Merchan-Solano, Rafael, & Ortiz-Perez, Hildelisa. (1987). Application of the regression method with dummy variables for the determination of norms for quantity of workers. *Economiay Desarrollo*, 100, 72-79.

Sawyer, Richard. (1986). Using demographic subgroup and dummy variable equations to predict college freshman grade average. *Journal of Educational Measurement*, 23, 131-145.

6.7.4 作为因果分析的路径分析

作为数据分析的一种形式,路径分析越来越普及,这是因为它为几组测量变量间因果关系的决定提供了可能。科学的主要目的是对社会现象建立理论的解释。Kaplan(1964)说过:

> 科学是对稳定性和不变性的探索。它也是一件确认在实践中被证明对控制未来实践非常重要的东西的工作。最基本的科学问题是"什么样的困难在这里徘徊?"(p.85)

当路径分析的潜在假设被满足时,在许多变量都被同时控制的情况下,理论和数据可能是相关的。路径分析本质上是在理论模型检验中使用标准多元回归方程的数据分析方法。

不应该或者至少不能对因果解释抱有过高的希望。无力处理社会系统中的所有变量,无力测量和规划他们精确的相互关系,这使得大部分问题的结果仅仅初步接近因果关系。然而方法的能力将继续挑战和激起研究者的兴趣,而且它的用处是增生扩散的。

一个研究者通常希望揭示自变量和因变量之间的关系。使用简单和复相关经常产生重要的关系,但它们从不证明因果关系。例如,如果我们希望父亲职业地位和儿子职业地位有关系,相关方法可以决定它们的相关关系,但是因果关系只能推断。使用路径分析,就可以假定作为自变量的父亲教育获得和职业地位是儿子后来的教育获得、首次职业地位和最近职业地位的原因。

(1)路径分析应用的6步骤

步骤1:建构一个因果图 路径分析允许社会理论家以线性因果模型的形式来陈述理论。重要的问题是不得不利用在因果或过程关系中的优先变量的顺序。因果模型包括一个社会现实意义上的过度简化模型的建构,且这种模型只考虑数量非常有限的、在特殊研究领域有重要作用的变量。寻找最重要的变量,所有其余变量都称为"剩余变量"。根据过去研究的结论和当前理论,社会科学家建构一个假定在变量中起作用的方法/过程的表达式。

假设我们利用阶层理论和研究。我们假定男性生命历程中地位变化象征着父亲教育获得(A)和父亲职业获得(B)将决定儿子后来的教育获得(X)、首次职业(Y)和

最近职业(Z)。这是线性陈述或者历时顺序,可以写成如下:

$$(A \to B) \to X \to Y \to Z$$

前面的变量可能不仅通过介入变量,而且直接地影响后面的变量。

步骤2:按顺序建立变量间的关系模式 概念上的框架必须转换成定量估测。这通过按顺序建立变量间的关系模式来完成。运用模型中的5个地位变量的简单相关做成相关矩阵。Blau和Duncan(1967)的修订显示了他们职业流动研究的矩阵(表6.3)。

表6.3 5个地位变量的简单相关

变量	Z	Y	X	B	A
Z 儿子的最近职业	—	0.541	0.596	0.405	0.322
Y 儿子的首次职业		—	0.538	0.417	0.332
X 儿子的教育			—	0.438	0.453
B 父亲的职业				—	0.516
A 父亲的教育					—

简单相关测量前置变量对后置变量的总的影响大小。最近职业是所有其他4个前置变量的期望结果。表6.3的第一行显示了所有4个前置变量对最近职业的重要相关,最高的是儿子的教育($r=0.596$),其次是他的首次职业($r=0.541$)。如期望的,父亲的职业和父亲的教育以递减大小相关(分别是$r=0.405$和$r=0.322$)。第二行呈现了与首次职业的相关。与父亲职业和教育也出现了同样的模式。第三行重复了儿子的教育与父亲职业和教育的期望关系。第四行证明了父亲的职业与教育的高度相关($r=0.516$)。

步骤3:描绘一个路径图 路径图一般地如图6.4所示,通过单向箭头来连接包括在基本模型中的一些或者所有变量。变量根据它们的理论排序从左到右分布。第一个自变量放在最左端。在这个例子里,自变量是父亲的教育和父亲的职业,且用双向箭头连接以区别于其他的影响路径。在模型中,不会被其他变量影响的变量(零阶相关,intercorrelations of zero order)称作外生变量(exogenous variables)。外生变量这一术语是指所有在模型之前和之外的变量。

剩余的变量子集(可能只包含一个变量)假定包括因变量,这些变量称作内生变量(endogenous variables,这里的X,Y和Z)。较外生变量而言,这个子集的变量被认为全部由系统中的一些变量组合决定。图6.4中的从一个测量变量到另一测量变量的直线表明了一个变量对另一个变量的直接影响。这里也存在间接影响,如在图中说明的那样。被认为受某一前置因素影响的变量,依次可能作为后置变量的原因。例如,X受A和B的影响,它反过来影响Y和Z。因此,Y和Z受A和B间接影响,还有一些直接影响。

最后,剩余路径必须画出。这些是没有来源标识的线条,箭头指向每个内生变量或者结果变量。剩余变量通过系统外的箭头指向X,Y和Z来表示。剩余变量是未被发现或未被测量的原因,是测量误差,与通过分析确定的真实关系的附加性、线性和特性相背离。

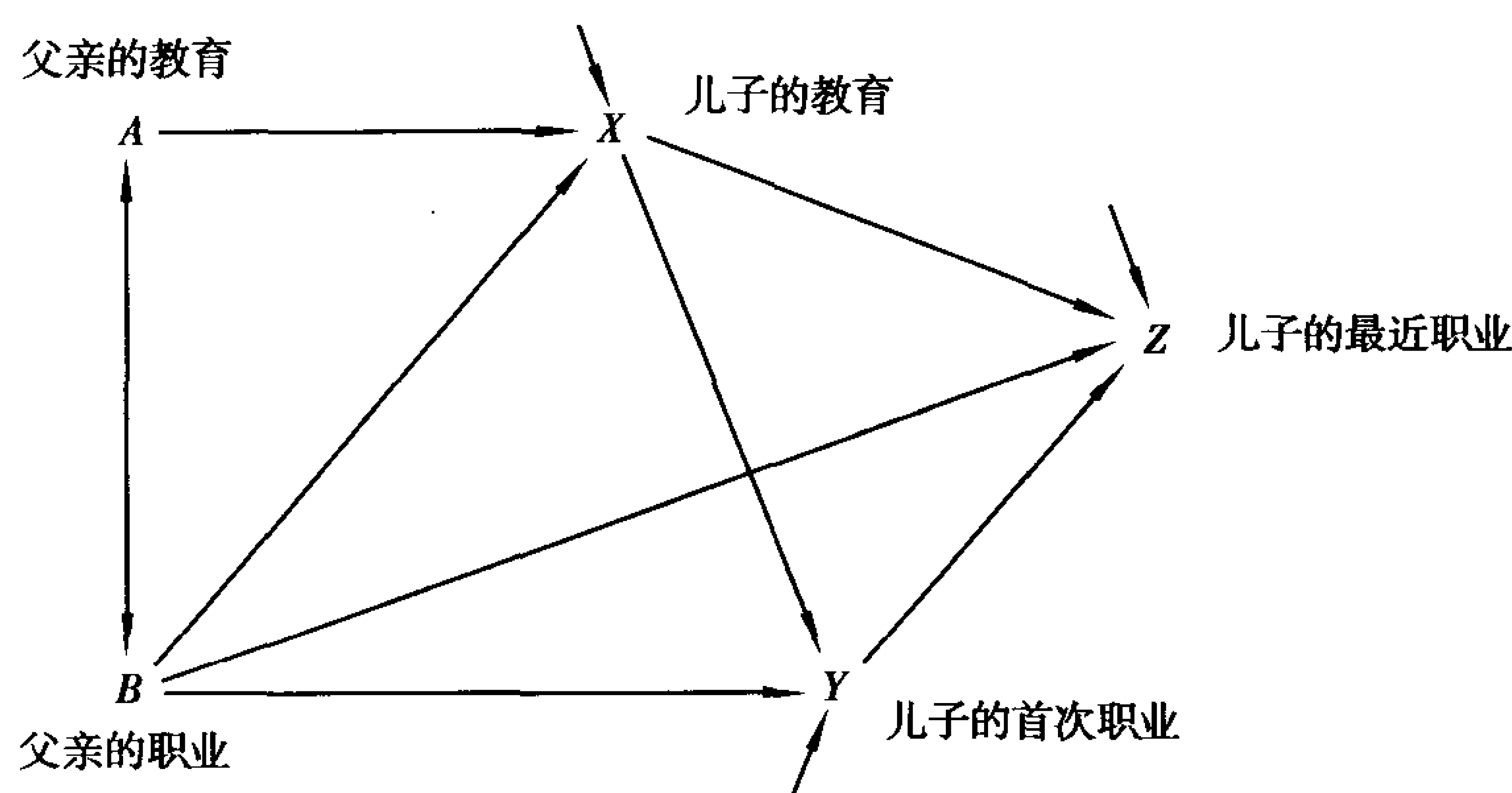


图 6.4 阶层形成过程的基本模型

资料来源：Adapted with permission of the Free Press, a Division of Simon & Schuster, Inc., from Peter M. Blau and Otis Dudley Duncan(1967), *The American Occupational Structure* (New York: John Wiley). Copyright 1967 by Peter M. Blau and Otis Dudley Duncan.

注：路径模型表现出的因果图的代数公式是建立在方程系统上的，而不是多元回归分析中经常使用的单个方程。这个特征允许推断影响的一个灵活排序。每条线代表一个调查和一个直接（或净）影响的因素。注意 Blau 和 Duncan 强调了父亲的职业作为一个原因因素——路径系数回溯到儿子的教育，首次职业和最近职业。从另一个方面，父亲的教育仅仅通过儿子的教育回溯。

步骤 4：计算路径系数 路径系数反映了当考虑其他相关变量影响的情况时，一个给定变量对另一个变量的直接贡献的大小。当变量以标准方式测量时，路径系数等同于偏回归系数(β)。常用两种计算路径系数的方法：第一种使用回归程序，即采用原始数据，并从标准化的数据输入中计算偏回归系数。路径系数和复相关系数一般都由标准的计算机回归程序提供[这个程序的解释，见 eg, Blau & Duncan(1967), pp. 171-177，或者 Loether & McTavish(1974), pp. 321-328]。第二种仅仅利用变量间的零阶相关，研究者可以使用“基本定理”来计算路径系数(Nygreen, 1971)。

图6.5 的路径系数已经标在路径图上，除了前置变量 A 和 B 以外。现在基本模型已经完成，只等检验了。

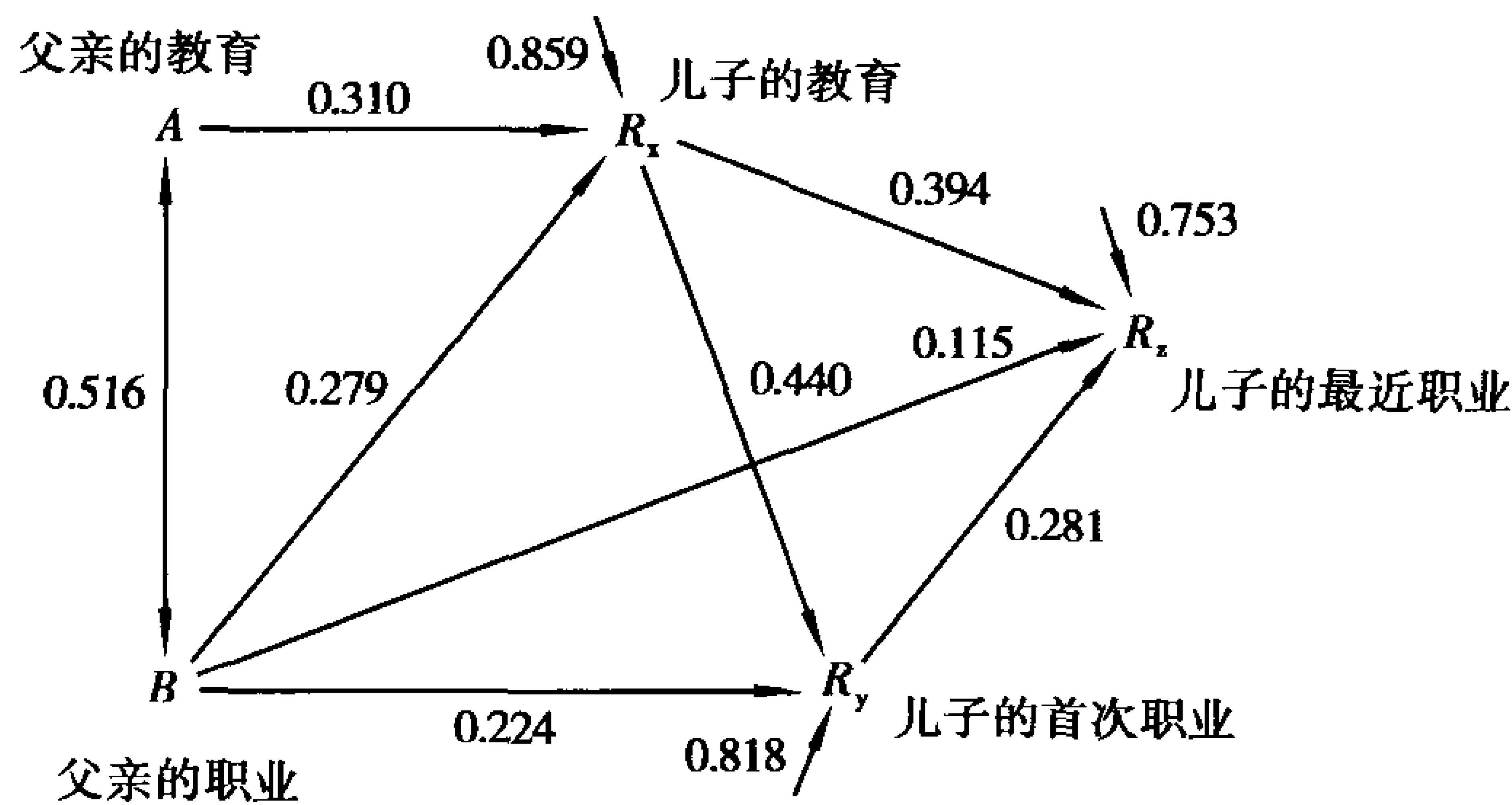


图 6.5 因果图的路径模型

步骤5:基本模型的拟合优度(goodness of fit)检验 分析的关键是观察到的数据与基本模型之间的拟合优度检验。通常可能有三种方法可以达到:

- 检查由模型中特定连接的变量所解释的因变量的误差总量。
- 检查路径系数的大小以发现它们是否足够大到保证变量或者路径存在于模型中。
- 评价模型预测相关系数的能力,它们(相关系数)不用于路径系数的计算。[进一步信息,请见 Land(1969)]。

一个调查者常常从这三个方面比较模型的有效性和备择模型的有效性。这是任何科学中解释进步的核心。

标准化的偏回归系数和变量具体组合的决定系数对于首次拟合优度标准至关重要。表6.4是 Blau 和 Duncan 数据的修订。它表明父亲的职业、父亲的教育和儿子的教育的决定系数是0.26,意味着儿子教育的26%的误差可以由父亲职业和教育来解释。

表6.4 标准化的偏回归系数(BETA)和变量具体组合的决定系数

因变量	首次职业	儿子的教育	父亲的职业	父亲的教育	决定系数(R^2)
儿子的教育			0.279	0.310	0.26
儿子的首次职业		0.433	0.214	0.026	0.33
儿子的最近职业	0.282	0.397	0.120	-0.014	0.43

同样地,儿子首次职业的33%的误差可以由父亲的职业、父亲的教育和儿子的教育解释。最后,儿子最近职业的43%的误差可以由父亲的职业、父亲的教育、儿子的教育和儿子的首次职业来解释。注意父亲的教育对解释这个误差是没有用的。未被解释的误差是 $1-R^2_{4.123}=0.57$。儿子最近职业模型还留有57%的误差没有解释。这个可能不如希望的那样令人满意。“未解释”的误差取决于模型中没有包括的变量或者测量误差。为了完全,$1-R^2$值被描述为剩余变量,如图6.5所标明的 R_X,R_Y和 R_Z。这些剩余路径很大,调查者必须重新检查他或她的因果图。然而,不应假定剩余变量的大小肯定是对成功解释所研究现象的一种测量。正如 Blau 和 Duncan(1967)所述,“与剩余变量相关的问题根本不是它的大小,而是它所代表的未被观察到的因素是否适当地被描绘成与所测量的前置因素不相关”(p. 175)。

按照拟合优度的第二个标准,模型是成功的。绝大部分路径系数都是显著的。这证明父亲的职业和儿子首次职业对父亲的教育的净回归(net regressions)如此小以至可以忽略。因此,父亲的教育可以被忽视,也不会丢失信息(Blau & Duncan, 1967, p. 175)。某人可能考虑消除父亲的教育这一因素,因为它的系数小,并重新估算路径系数。

第三,可以检验以前不用公式计算路径系数所观察的相关与从模型中获得的相关系数的预测之间的“拟合”。在这种情况下,父亲的教育与儿子首次职业(也与儿子最近职业)之间的相关不用于评估路径系数。

检验一个备择模型的问题包括基本模型的彻底重试。有两种可能:①模型中的替代因素被赋予更大的重要性;②增加新因素到模型中。不论路径图或者因果图的描绘

被判定为合适与否都依赖于理论与经验的考虑。在以下这一意义上因果图式必须完整,即所有解释的原因,以及假定与因变量不相关的未测原因都必须被提出。

步骤6:解释结果 研究因果图中的变量的直接和间接效果。父亲的职业对儿子的教育、首次职业和最近职业的直接效果由路径系数0.279,0.224和0.115表示,且没有一个特别大,间接效果的累积仍然都显著。父亲的职业和教育确实影响儿子的教育,且这依次影响儿子的首次职业,儿子的首次职业依次影响儿子的最近职业。同时,甚至更大影响的许多其他因素明显地正决定这个最后影响的因变量。

路径分析的方法不是发现因果规律的一种方法,而是寻求对既定总体中运作的假定因果系统作定量解释的一个程序。

参考文献

Blau, Peter M., & Duncan, Otis Dudley. (1967). *The American occupational structure.* New York: John Wiley.

Land, Kenneth C. (1969). Principles of path analysis. In Edgar F. Borgatta (Ed.), *Sociological methodology* 1969, pp. 1-37. San Francisco: Jossey-Bass.

Loether, Herman J., & McTavish, Donald C. (1974). *Descriptive statistics for sociologists.* Boston: Allyn & Bacon.

Nygreen, G. T (1971). Interactive path analysis. *American Sociologist*, 6, 37-43.

关于路径分析的推荐读物

Bishop, R. M., & Bieschke, K. J. (1998). Applying social cognitive theory to interest in research among counseling psychology doctoral students: A path analysis. *Journal of Counseling Psychology*, 45, 182-188.

Blalock, Hubert M., Jr. (1964). *Causal inferences in nonexperimental research.* Chapel Hill: University of North Carolina Press.

Boudon, Raymond. (1965). A method of linear causal analysis: Dependence analysis. *American Sociological Review*, 30, 365-374.

Cnaan, Ram A., Hasenfeld, Yeheskel, Cnaan, Avital, & Rafferty, Jane. (1993). Cross-cultural comparison of attitudes toward welfare-state programs: Path analysis with log-linear models. *Social Indicators Research*, 29, 123-152.

Costner, Herbert L., & Leik, Robert K. (1964). Deductions from "axiomatic theory." *American Sociological Review*, 29, 819-835.

Dellva, Wilfred L., Teas, R. Kenneth, & McElroy, James C. (1985). Leader behavior and subordinate role stress: A path analysis. *Journal of Political and Military Sociology*, 13,183-193.

Ensminger, Margaret E., & Slusarcick, Anita L. (1992). Paths to high school graduation or drop-out: A longitudinal study of a first-grade cohort. *Sociology of Education*, 65, 95-113.

Forbes, H. D., & Tufte, E. R. (1968). A note of caution in causal modelling. *American Political Science Review*, 62, 1258-1264.

Ganzeboom, Harry B. G., Treiman, Donald J., & Ultee, Wout C. (1991). Comparative intergenerational stratification research: Three generations and beyond. *Annual Review of Sociology*, 17,277-302.

Haney, Colleen J., & Long, Bonita C. (1995). Coping effectiveness: A path analysis of self-efficacy, control, coping, and performance in sport competitions. *Journal of Applied Social Psychology*, 25,17-26.

Heise, D. R. (1969). Problems in path analysis and causal inference. In Edgar F. Borgatta (Ed.), *Sociological methodology* 1969 (pp. 38-71). San Francisco: Jossey-Bass.

Hemmings, Brian, Putai, Jin, & Low, Renae. (1997). Evaluating a theoretical model of school withdrawal using a qualitative analysis. *Asia Pacific Journal of Social Work*, 7, 77-96.

Holland, Paul W. (1988). Causal inference, path analysis, and recursive structural equations models. *Sociological Methodology*, 18,449-484.

Hutchison, Steven. (1997). A path model of perceived organizational support. *Journal of Social Behavior and Personality*, 12, 159-174.

Israels, A. Z. (1987). Path analysis for mixed qualitative and quantitative variables. *Quality and Quantity*, 21, 91-102.

Krymkowski, Daniel H. (1991). The process of status attainment among men in Poland, the U.S., and West Germany. *American Sociological Review*, 56, 46-59.

Lee, Valerie E., & Frank, Kenneth A. (1990). Students' characteristics that facilitate the transfer from two-year to four-year colleges. *Sociology Education*, 63,178-193.

Lewis, Michael. (1999). A path analysis of the effect of welfare on infant mortality. *Journal of Sociology and Social Welfare*, 26, 125.

Li, C. C. (1955). *Population genetics.* Chicago: University of Chicago Press.

Loehle, Craig. (1994). A critical path analysis of scientific productivity. *Journal of Creative Behavior*, 28, 33-47.

McDonald, Roderick P. (1994). The bilevel reticular action model for path analysis with latent variable. *Sociological Methods and Research*, 22,399-413.

Simon, Herbert A. (1957). *Models of man.* New York: John Wiley.

Stinchcombe, Arthur. (1968). *Construction of social theories.* NewYork: Harcourt, Brace &World.

6.7.5 因子分析:以简单术语解释众多变量间的关系

在此,我们的目的是给社会科学的研究者提供因子分析基本概念的工作知识,而不是因子分析的统计细节。这一点将在以下的讨论中得到确定。但是,使用者需要掌握一些相关和回归系数的意义。

因子分析是一个研究概率的程序,很多变量中存在可以理解它们之间相互关系的共同因子的概率很小。正如 Schuessler(1971)所解释的,"我们观察到,在阅读获得高分的学生倾向于在拼写和算术也得高分。我们把学生分数中的连贯性或者相关归因于智力这一普通因素。"(p. 44)这个原理坚称,如果在其他方面没有共同之处,一系列绝对相同的事件存在的共有的情境,可能便是这个事件的原因。因此,为了揭示一个事件的原因,我们搜索那个经常当事件发生时出现的惟一的情境。在相似的意义上,通过因子分析,我们尝试分离出那些从两个或者多个变量中提出的共同要素,并且将这些变量间的相关归因于这一共同要素。可见因子分析是一个决定许多变量中的相关能否得出一些共同因子的数学程序。

C 可能被认为是 X 和 Y 的原因，或者仅仅是两个变量中出现的一个要素（图 6.6），因子分析认为 X 和 Y 的概率是同一件事的指示物。从观察到的 X 和 Y 之间的相关，可以做出它们由相同的原因引起，或者它们不同程度上处于相同事情的不同方面的推论。

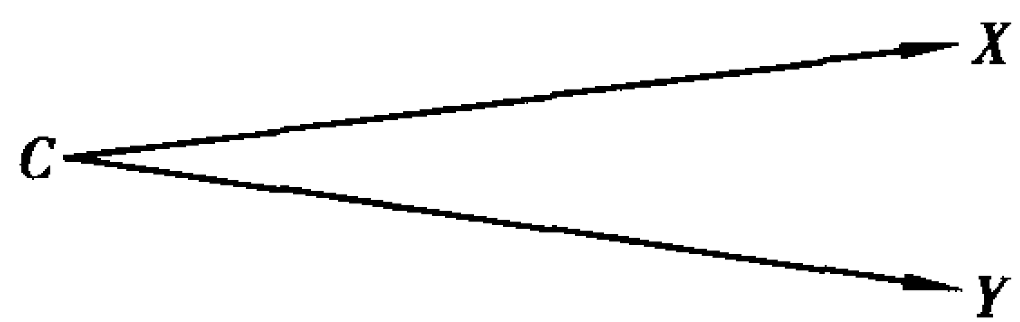


图 6.6　原因模型

（1）潜在的因子

为了区别所观察到的变量（操作的）和共同变量（藏在一般变量中），通常称后者为因子而不是变量。因此，最简单的一个例子，两个变量 Z_1 和 Z_2，二者都由两部分组成，一部分为二者所共有（A），另一个部分为二者各自所独有（B_i），那么：

$$Z_1 = A + B_1$$
$$Z_2 = A + B_2$$

按照惯例，Z 被称为一个变量，A 和 B_i 为因子。在深入进行之前，必须强调在所有方面的通常意义上，因子都是统计的变量：因子拥有均值和方差，它们可能对称分布，它们也可能与其他因子相关。“因子”这个专门术语有助于维持复合变量（观察到的）与构成部分（假设的）之间的区别。

参考文献

Schuessler, Karl. (1971). *Analyzing social data.* Boston: Houghton Mifflin.

（2）因子分析的简化能力

因子分析的惟一最与众不同的特征是它的数据简化能力。给出一组变量的一批相关系数，因子分析的方法帮助研究者发现是否存在一些关系的潜在模式，以使数据可能被“重新安排”或者简化到一组因子，这组因子在数量上比原来的那组变量少，这组因子可以被视为解释数据中所观察到的相互关系的原初变量。这种统计能力具有多种用法，但是这种方法最常用的是下列三种分类中的一种：①探测性（explortory）的用法，变量模式的探索和发现，以新概念的发现和数据的可能简化为目的；②证实性（confirmtory）的用法，关于根据显著因子和因子负荷的期望数量的变量结构的假设检验；③作为测量的方法使用，索引的构成作为新的变量用于以后的分析（Nie, Bent, & Hull, 1970）。

参考文献

Nie, N. H., Bent, D. H., & Hull, C. H. (1970). Statistical Package for the social Scieuces. NY: McGraw-Hill.

(3)作为研究设计的因子分析

仅有少数几种方法能够通过在不可操作的情境中的操作测试出未来的走势，因子分析是其中的方法之一。它是通过统计的方法，而不是操作控制的、传统的实验路线来寻求结论。

(4)因子分析的三个主要步骤

因子分析包括相当多种统计的方法，但基本上因子分析程序有三个步骤：

1)准备一个相关矩阵。

2)提取最初的因子——可能的数据简化的探测。

3)旋转出最终的结果——寻找简单能解释的因子。

这三个阶段中，每一阶段的主要选择可以由三个两分法来概括：步骤1中的 R 类型因子分析对 Q 类型因子分析，步骤2中解释因子对推断因子，步骤3中的正交旋转法对斜交旋转法。这些两分法在以下所列出的参考书目中的因子分析解决办法中有详细的解释。

关于因素分析的推荐读物

Attias, H. (1999). Independent factor analysis. *Neural Computation*, 11, 803-851.

Cattell, Raymond B. (1952). *Factor analysis: An introduction and manual for the psychologist and social scientist*, New York: Harper & Row.

Crabtree, Craig. (1999). Critical factor analysis: An exploratory general model for understanding diverse systems. *World Futures*, 53, 213-227.

Dunlap, William P., & Landis, Ronald S. (1998). Interpretations of multiple regression borrowed from factor analysis and canonical correlation. *Journal of General Psychology*, 125, 397-407.

Fruchter, Benjamin. (1954). *Introduction to factor analysis.* Princeton, NJ: Van Nostrand.

Harman, Harry H. (1976). *Modern factor analysis* (3rd ed.). Chicago: University of Chicago Press.

Heck, Ronald H. (1998). Factor analysis: Exploratory and confirmatory approaches. In George Marcoulides (Ed.), *Modern methods for business research: Methodology for business and management* (pp. 177-215). Mahwah, NJ: Lawrence Erlbaum.

Hershberger, Scott L. (1998). Dynamic factor analysis. In George Marcoulides (Ed.), *Modern methods for business research: Methodology for business and management* (pp. 217-249). Mahwah, NJ: Lawrence Erlbaum.

Horst, Paul. (1965). *Factor analysis of data matrices.* New York: Holt, Rinehart & Winston.

Hutchinson, Susan R. (1998). The stability of post hoc model modifications in confirmatory factor analysis models. *Journal of Experimental Education*, 66, 361-380.

Jöreskog, K. G. (1979). *Advances in factor analysis and structural equation models.* Cambridge, MA: Abt.

Kim, Jae-on, & Mueller, Charles W. (1978). *Factor analysis: Statistical methods and practical issues.* Beverly Hills, CA: Sage.

MacCallum, Robert C., Widaman, Keith F., Zhang, Shaobo, & Hong, Sehee. (1999). Sample size in factor analysis. *Psychological Methods*, 4, 84-99.

Marsh, Herbert W., Hau, Kit Tai, Balla, John R., & Grayson, David. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research*, 33, 181-220.

Nie, N. H., Bent, D. H., & Hull, C. H. (1970). *Statistical package for the social sciences.* New York: McGraw-Hill.

Raykov, Tenko. (1998). On the use of confirmatory factor analysis in personality research. *Personality and Individual Differences*, 24, 291-293.

Raykov, Tenko, & Little, Todd D. (1999). A note on procrustean rotation in exploratory factor analysis: A computer intensive approach to goodness-of-fit evaluation. *Educational and Psychology Measurement*, 59, 47-57.

Rummel, Rudolph J. (1970). *Applied factor analysis.* Evanston, IL: Northwestern University Press.

Thurstone, Louis L. (1947). *Multiple factor analysis.* Chicago: University of Chicago Press.

因素分析研究范例

Kalliath, Thomas J., Bluedorn, Allen C., & Gillespie, David F. (1999). A confirmatory factor analysis of the competing values instrument. *Educational and Psychology Measurement*, 59, 143-158.

Kranzler, John H., & Keith, Timothy Z. (1999). Independent confirmatory factor analysis of the Cognitive Assessment System. *School Psychology Review*, 28, 117-144.

Owens, Timothy J. (1993). Accentuate the positive and the negative: Rethinking the use of selfesteem. *Social Psychology Quarterly*, 56, 288-299.

Shevlin, M. E., & Lewis, Christopher Alan. (1999). The Revised Social Anxiety Scale: Exploratory and confirmatory factor analysis. *Journal of Social Psychology*, 139, 250-252.

Valtinson, Gale Rene. (1998). A multi-sample confirmatory factor analysis of work-family conflict. *Dissertation Abstracts International, Section B: The Sciences and Engineering.*

6.8 其他统计方法的概述及它们应用的例子

这节里的概念通常出现在主要杂志上的社会学研究论文中。此处的解释尝试尽可能简单地陈述这些概念的意义。每个概念讨论之后给出的参考书目指出了研究的应用。这节的目标是给读者提供理解而不是解释和阐明操作的程序。

6.8.1 Bayesian 方法:贝氏推论和贝氏统计

Bayesian(贝叶斯,简称贝氏)认为科学家在进行他们实验之前必须具体化他们的

意见，形成推测概率，做实验就是收集有关这些意见的数据，接着正式使用贝氏定理修正前面的概率，形成新的在后的概率。在后的概率是已经根据数据提供的信息修正过的科学家的意见。这是所有贝氏方法背后的关键思想。

统计上讲，贝氏推论仅仅是一种条件概率。如果结果（证据）已经出现，贝氏推论提供原因（假设）的概率。推论的通常顺序被颠倒了：从结果到原因而不是从原因到结果。在贝氏推论对主观意见的强调上，它具有提炼某种概率的与众不同的特征。

贝氏方法很少使用于社会学，但是它们已经在经济学和心理学领域得到积极应用。正如分析归纳方法所显示的那样，无论何时，社会分析家从结果到原因反向工作，贝氏方法似乎对社会学都是有价值。在分析归纳中，无论何时提出结果，都必须提出原因，且无论何时出现结果，原因也必须出现。像贝氏统计这种统计方法，利用参数在前的信息（客观或主观）。借助于主观或个人概率的概念，贝氏统计术语也已经被应用于统计方法中。

关于贝氏方法的推荐读物

Arminger, Gerhard, & Muthen, Bengt O. (1998). A Bayesian approach to nonlinear latent variable, models using the Gibbs Sampler and the Metropolis-Hastings. *Psychometrika*, 63, 271-300.

Bauwens, Luc., Lubrano, Michele, & Richard, Jean Francois. (2000). *Bayesian inference in dynamic econometric models.* New York: Oxford University Press.

Djafari, Ali M. (Ed.). (1998). *Bayesian inferences for inverse problems.* Bellingham, WA: SPIE International Society for Optical Engineering.

Gross, Alan L. (1997). Interval estimation of bivariate correlations with missing data on both variables: A Bayesian approach. *Journal of Educational and Behavioral Statistics*, 22, 407-424.

Kontsevich, Leonid L., & Tyler, Christopher W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research*, 39, 2729-2737.

Lee, Peter M. (1997). *Bayesian statistics: An introduction.* New York: Oxford University Press.

Park, Young C., & Choi, Key Sun. (1996). Automatic thesaurus construction using Bayesian networks. *Information Processing and Management*, 32, 543-553.

van der Linden, Wim J. (1998). Bayesian item selection criteria for adaptive testing. *Psychometrika*, 63, 201-216.

Winkler, Robert L. (1993). Bayesian statistics: An overview. In Gideon Keren & Charles Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Statistical issues* (pp. 201-232). Hillsdale, NJ: Lawrence Erlbaum.

Zellner, Arnold. (1996). *An introduction to Bayesian inference in econometrics.* New York: John Wiley and Sons.

6.8.2 典型相关

典型相关是两组自变量和因变量之间的最大相关。它可以与被比作复相关的简单系数相比较，复相关提供的是许多自变量和一个因变量之间的最大相关。

典型相关的基本思想是，通过最小平方分析，形成两个线性组合，一组是自变量，X_J，一组是因变量，Y_N。这两个组合间的相关就是典型相关 R_c。典型相关的平方，R_c^2，

是对这两组组合共享方差的估计。

关于典型相关的推荐读物

Alexander, Erika D. (2000). *Using canonical correlation to explore relationships between sets of variables: An applied example with interpretive suggestions.* Paper presented at the annual meeting of the Southwest Educational Research Association, Dallas, TX.

Arnold, Margery E. (1996). *The relationship of canonical correlation analysis to other parametric methods.* Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans, LA.

Cook, J. A., Razzano, L., & Cappelleri, S. (1996). Canonical correlation analysis of residential and vocational outcomes following psychiatric rehabilitation. *Evaluation and Program Planning*, 19, 351-363.

Fan, Xitao. (1997). Canonical correlation and structural equation modeling: What do they have in common? *Structural Equation Modeling*, 4, 65-79.

Fan, Xitao, & Wang, Lin. (1996). Comparability of jackknife and bootstrap results: An investigation for a case of canonical correlation analysis. *Journal of Experimental Education*, 64, 173-189.

6.8.3 因果分析

因果分析在所有社会科学分支学科中被赋予了极大的优先地位。因果判断被用来解释事件的发生，理解为什么特别的事件会发生。拥有因果知识，预测事件和运用一些控制它们的方法经常是可能的。关于这个主题，Marini 和 Singer(1988)提供了一个极好的参考书目。

Berk(1988)提供了一个极好的社会学因果推论的评论。他指出，在实验方法不可用时，Hubert M. Blalock 和 Otis Duncan 提出将“结构方程”(structaral equation)来作为从中可以得出因果推论的模型，增加的注意力都集中在因果关系上。Smelser(1988)在文集的第三部分把路径分析描述为因果分析，指出了其前景和局限性。Bernett(1983)，在他 20 世纪 70 年的美国社会学的因果术语学调查中，已经观察到在社会学里，因果关系已经被证明为是一个脆弱的(fragile)的核心概念。

参考文献

Berk, Richard A. (1988). Causal inference for sociological data. In Neil J. Smelser (Ed.), *Handbook of sociology* (pp. 155-171). Newbury Park, CA: Sage.

Bernett, Christopher. (1983). The career of causal analysis in American sociology. *British Journal of Sociology*, 34(2), 230-254.

Marini, Margaret M., & Singer, Burton. (1988). Causality in the social sciences. In Clifford C. Clogg (Ed.), *Sociological methodology* 1988 (pp. 347-409). Washington, DC: American Sociological Association.

Smelser, Neil J. (Ed.). (1988). *Handbook of sociology.* Newbury Park, CA: Sage.

关于因果分析的推荐读物

Finkel, Steven E. (1995). *Causal analysis with panel data.* Thousand Oaks, CA: Sage.

Kish, Leslie. (1987). *Statistical design for research.* New York: John Wiley.

Levine, Tamar, & Donitsa-Schmidt, Smadar. (1998). Computer use, confidence, attitudes, and knowledge: A causal analysis. *Computers in Human Behavior*, 14, 125-146.

McClendon, McKee J. (1994). *Multiple regression causal analysis.* Itasca, IL: F. E. Peacock.

Okraku, Ishmael O. (1998). Tavern-going in America: A causal analysis. *Leisure Sciences*, 20, 303-317.

Peyrot, Mark. (1996). Causal analysis: Theory and application. *Journal of Pediatric Psychology*, 21, 3-24.

Valette-Florence, Pierre. (1998). A causal analysis of means-end hierarchies in a cross-cultural context: Methodological refinements. *Journal of Business Research*, 42, 161-166.

Xiaoming, Li, Feigelman, Susan, Stanton, Bonita, Galbraith, Jennifer, & Huang, Weihua. (1998). Drug trafficking and drug use among urban African-American adolescents: A causal analysis. *Journal of Adolescent Health*, 23,280-288.

Yates,Donald L., & Pillai, Vijayan K. (1996). Attitudes toward community policing: A causal analysis. *Social Science Journal*,33,193-209.

6.8.4　聚类(cluster)分析

聚类分析是研究者依据项目间的相互关系显示把项目划分成“一群”的一个方法。例如,我们知道人们选择工作时考虑的原因有:创新的能力,与人合作的机会,获得足够数目金钱和累积财产的欲望。如果关于工作特征的各种问题都涌向回答者,研究者就可能发现项目将分成三个普通的类别。较类别外的其他项目而言,每一个类别将包括在它们的类别里和其他项目有着更高相互关系的项目。

相关矩阵可以通过矩阵重新排列的行和列来完成,由此,沿着对角线的不同点上三角排列成的“串”的项目就有更高相互关系。通过这个方法,相互作用力的本质可以被精确地揭示出来。

关于聚类分析的推荐读物

Bruehl, Stephen, Lofland, Kenneth R., Semenchuk, Elizabeth M., Rokicki, Lori A., & Penzien, Donald B. (1999). Use of cluster analysis to validate HIS diagnostic criteria for migraine and tension-type headache. *Headache*, 39, 181-189.

Fleming, Jennifer M., Strong, Jenny, & Ashton, Roderick. (1998). Cluster analysis of self-awareness levels in adults with traumatic brain injury and relationship to outcome. *Journal of Head Trauma Rehabilitation*, 13, 39-51.

Prior, Margot, Eisenmajer, Richard, Leekam, Susan, Wing, Lorna, Gould, Judith, Ong Ben, & Dowe, David. (1998). Are there subgroups within the autistic spectrum? A cluster analysis of a group of children with autistic spectrum disorders. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 39, 893-902.

Roussos, Louis A., Stout, William F., & Marden, John I. (1998). Using new proximity measures with hierarchical cluster analysis to detect multidimensionality. *Journal of Educational Measurement*, 35, 1-30.

Seifert, Timothy L. (1997). Academic goals and emotions: Results of a structural equation model and a cluster analysis. *British Journal of Educational Psychology*, 67, 323-338.

6.8.5 同期群(cohort)分析

术语"同期群"指的是通过年龄或者其他选择的特征把人进行匹配或者近视匹配,由此所定义的特征可以得到持续的支持。同期群分析是对指定变量所评估的一个或者多个同期群的研究。对职业和工业社会学而言,这种研究是非常有价值的,这是因为劳动力中的同期群都是通过代表性的和纵贯的设计来研究的。人口学、健康科学、犯罪学、老人医学和许多其他领域的学科经常使用同期群分析。

关于同期群分析的推荐读物

Birkett, Nicholas J. (1997). Trends in smoking by birth cohort for births between 1940 and 1975: A reconstructed cohort analysis of the 1900 Ontario Health Survey. *Preventive Medicine*, 26 534-541.

Glenn, Norval D. (1977). *Cohort analysis.* Beverly Hills, CA: Sage.

Greenberg, David F., & Larkin, Nancy J. (1996). Age-cohort analysis of arrest rates. In David F. Greenberg (Ed.), *Criminal careers* (Vol. 1, pp. 91-104). Aldershot, UK: Dartmouth.

Miller, Alan S., & Nakamura, Takashi. (1997). Trends in American public opinion: A cohort analysis of shifting attitudes from 1972-1990. *Behaviormetrika*, 24, 179-191.

Riley, Matilda W. (1973). Aging and cohort succession: Interpretations and misinterpretations. *Public Opinion Quarterly*, 37, 35-49.

Takei, Noriyoshi, Lewis, Glyn, Sham, Pak C., & Murray, Robin M. (1996). Age-period-cohort analysis of the incidence of schizophrenia in Scotland. *Psychological Medicine*, 26, 963-973.

Yoon, In Jin. (1997). A cohort analysis of Korean immigrants' class backgrounds and socioeconomic status in the United States. *Korea Journal of Population and Development*, 26, 61-81.

6.8.6 沟通网络分析

这是确认系统中沟通网络研究的一个方法。许多人类行为发生在相互作用中,通过相互作用一个个体和另一个或者更多其他个体交换信息。一个沟通网络描绘出由模式的信息流所连接起来的个体。分析包括各种研究程序:派系的鉴别,专门沟通角色(联络、连接、隔离)的检查和信息结构化索引的建构,例如个体、两分体、人际网络、派系或者整个系统的信息连通性。

6.8.7 因变量的分解

许多社会变量都是合成物。举个例子,人口增长是自然增长和净移民的总和。其中的每一个又可以进一步分解:自然增长可以用出生数减去死亡数计算,净移民是移入与移出之间的差额。当这种分解是可能的时,有利于:①计算在合成变量中成分对

变差的相关贡献;②确定原因影响合成变量是如何通过各自的成分传递的。

6.8.8 因果模型中效果的分解

解释路径模型和其他结构方程模型中直接和间接原因模式是很重要的。这种解释帮助回答以下问题:

- 变量 X 如何影响变量 Y?
- 机制 Z 对 X 或者 Y 影响的贡献有多大?
- 机制 Z 对解释总体 A 中 X 对 Y 影响的贡献与总体 B 中是否一样大?

因果模型经常把前置变量的总体效果分解成直接的和间接的成分。一个变量对另一个变量的直接影响只是不通过介入变量传递的总体效果中一部分。间接影响是由介入模型中影响的原因与效果之间的变量传递或者传达的一个变量的总体效果的那个部分。这里的分解意味着计算多元方程模型中总体、直接和间接的效果。

关于因果模型中效果分解的推荐读物

Alwin, Duane F., & Hauser, Robert M. (1975). The decomposition of effects in pattern analysis. *American Sociological Review*, 40,37-47.

6.8.9 特尔斐技术

特尔斐技术是一个产生专家对任何一个特定主题意见的预测方法。它是在 20 世纪 50 年代初期,由 RAND 公司在试图消除专家参与会议时人际间的情感和行为所引起的影响的控制变量时发展的。特尔斐技术使用书面的答案而不是把专家放在一个面对面的讨论会上。匿名性避免了一个群体会议中特定个体的可能支配。当专家分散广泛且不易亲自见面时,这个方法特别有用。如果回答者对另一个人有敌意或者对抗,这个方法也可以用。

关于特尔斐技术的推荐读物

Martin, Andrew G., & Frick, Martin J. (1998). The Delphi technique: An informal history of its use in agricultural education research since 1984. *Journal of Agricultural Education*, 39, 73-79.

Passig, David. (1993). Reactions to experts' forecasts by a group of Jewish teenagers: An Imen-Delphi exercise—An applied social methodology-a variant of the Delphi-forecasting technique: I and II. *Dissertation Abstracts International*, 54, 480.

Petrina, Stephen, & Volk, Kenneth S. (1992). Policy making processes and the Delphi technique in STS curricula: A case study examining energy issues. *Bulletin of Science, Technology and Society*, 12,299-303.

Stahl, Nancy N., & Stahl, Robert J. (1991). We can agree after all! Achieving consensus for a critical thinking component of a gifted program using the Delphi technique. *Roeper Review*, 14, 79-88.

Wicklein, Robert C. (1993). Identifying critical issues and problems in technology education using a modified-Delphi technique. *Journal of Technology Education*, 5, 54-71.

6.8.10 判别(函数)分析

判别函数最初是由 R. A. Fisher 提出的。它被设计来帮助将一个独立的观察值分成两个组中的一个。判别函数已经被定义为一组 n 变量的一个线性组合,这组 n 变量将被分成两个不同的种类(或者组),且 n 变量测量的事件或者项目是可行的,而且也将对误分类最小可能部分有用。举个例子,在将人分成两个社会群体的问题中,比如,文化吸引和文化排斥,判别分析是有用的。

关于判别(函数)分析的推荐读物

D'Andrea, Livia M., & D'Andrea, Lester R. (1996). Prediction of treatment outcome using state-trait anxiety inventory scores: The use of discriminant analysis. *Journal of Addictions & Offender Counseling*, 16, 50-61.

Ermer, Julie, & Dunn, Winnie. (1998). The Sensory Profile: A discriminant analysis of children with and without disabilities. *American Journal of Occupational Therapy*, 52,283-290.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7, 179-188.

Huberty, Carl J., & Lowman, Laureen L. (1997). Discriminant analysis via statistical packages. *Educational and Psychological Measurement*, 57, 759-784.

Lagana, Luciana. (1995). Older adults' expectations about mental health counseling: A multivariate and discriminant analysis. *International Journal of Aging & Human Development*, 40,297-316.

Pineda, David, Ardila, Alfredo, & Rosselli, Monica. (1999). Neuropsychological and behavioral assessment of ADHD in seven-to twelve-year-old children: A discriminant analysis. *Journal of Learning Disabilities*, 32, 159-173.

6.8.11 事件史分析

事件史分析运用方法来分析关于数量、时间以及人与实体经历的事件序列的数据。本质上,一个事件是一个状态的变化,就像从单身到结婚的变化。事件由时间间隔来隔离。婚姻作为一个事件直到离婚或者死亡发生时结束。事件也可以用于对像经历公司倒闭,国家战争、沦陷和革命的公司或者国家这种实体进行分析。

事件分析对社会学有很大的潜在用处。已经形成的研究程序运用事件发生顺序的相关数据而非精确的日期数据进行分析。社会学家和经济学家已经发现事件分析在工作历史和各种劳动力动力学问题研究中的用处。顺序数据可用于对年生产力、整合学校里孩子的数量、每年由犯罪造成的损失等研究之中。

人与实体都适用的普遍分析方法涉及一个模型的提出,该模型将国家不发生变迁的可能性,国家变迁的可能性或者不同国家间运动的可能性联系在一起。事件史分析的一个例子是被用于评估人口参数。

关于事件史分析的推荐读物

Allison, Paul D. (1985). *Event history analysis: Regression for longitudinal event data.* Beverly Hills, CA: Sage.

Fernandez, Elizabeth. (1999). Pathways in substitute care: Representation of placement careers of children using event history analysis. *Children & Youth Services Review*, 21, 177-216.

Leiter, Jeffrey, & Johnsen, Matthew C. (1997). Child maltreatment and school performance declines: An event-history analysis. *American Educational Research Journal*, 34, 563-589.

Myers, Daniel J. (1997). Racial rioting in the 1960s: An event history analysis of local conditions. *American Sociological Review*, 62, 94-112.

Palmer, William R. T. (1997). Parole prediction using current psychological and behavioural predictors, a longitudinal criterion, and event history analysis. *Dissertation Abstracts International*, 58, 2694B.

6.8.12 对数线性模型与分析

对数线性模型,相依表(交互表,contingency table)分析的最近发展,将新的方法应用于分析多元表格。在过去,纯粹的分类数据难以分析,特别在多变量模型的建构中更是如此。Leo A. Goodman 说过,对数线性模型不仅帮助人发现双变量间的关系和复杂表格里的高阶相互关系,而且也可以看作类似于联立方程分析(analysis of simultaneous equation)。其在民意调查及其他调查方面的探索可能已经赶上因果和路径分析。一些研究者警告说,应用于定距层次数据的同样概念不能清楚地用于解释分类数据的交叉分类。

对数线性模型的本质是为由所谓的主效应和交互效应引介的多元总体交叉分类的期望频数建构一个模型。有时,它更便于对期望频数的自然对数起作用,使用于评估在模型下的期望频数的一组“拟合边际”(fitted marginals)表格将描绘出一个模型。

关于对数线性模型与分析的推荐读物

Keller, F., & Wolfersdorf, M. (1995). Changes in suicide numbers in psychiatric hospitals: An analysis using log-linear time-trend models. *Social Psychiatry and Psychiatric Epidemiology*, 30, 269-273.

Knoke, David, & Burke, Peter J. (1980). *Log-linear models.* Beverly Hills, CA: Sage.

Miller, Todd Q., & Flay, Brian R. (1996). Using log-linear models for longitudinal data to test alternative explanations for stage-like phenomena: An example from research on adolescent substance abuse. *Multivariate Behavioral Research*, 31, 169-196.

Rojewski, Jay W., & Bakeman, Roger. (1997). Log-linear modeling and analysis: Reflections on the use of multivariate categorical data in social science research. *Exceptionality*, 7, 199-203.

Sobel, Michael E. (1998). Some log-linear and log-nonlinear models for ordinal scales with midpoints, with an application to public opinion data. *Sociological Methodology*, 28, 263-292.

Vermunt, Jeroen K. (1997). *Log-linear models for event histories.* Thousand Oaks, CA: Sage.

6.8.13 马尔可夫链(markov chains)

马尔可夫链是预测随时间的过去而发生的变化的模型。设想一个预测随时间过去的结果的问题,一些人从一个工作流动到另一个工作,然而其他人却仍然停留于同样的工作。当面临这样一个问题和相关的数据时,想知道观察频数是否可以由简单概率规律说明是自然的。马尔可夫规律说明一个固定的,但不必相等的,朝向时间序列结束的流动者和滞留者的分类。一个马尔可夫链可能因开始于两分术语中的各种变化行为而被发现:那些在选举日前改变他们的想法和那些从未改变他们想法的,那些士气每天变化和那些士气保持不变的,等等。

关于马尔克夫链的推荐读物

Donhardt, G. L. (1995). Tracking student enrollments using the Markov chain, comprehensive tool for enrollment management. *Journal of College Student Development*, 36, 457-462.

Humphreys, Keith. (1998). The latent Markov chain with multivariate random effects: An evaluation of instruments measuring labor market status in the British Household Panel Study. *Sociological Methods and Research*, 26, 269-299.

Meiser, Thorsten, & Ohrt, Barbara. (1996). Modeling structure and chance in transitions: Mixed latent partial Markov-chain models. *Journal of Educational and Behavioral Statistics*, 21,91-109.

Yamaguchi, Kazuo. (1996). Some log-linear fixed-effect latent-trait Markov-chain models: A dynamic analysis of personal efficacy under the influence of divorce/widowhood. *Sociological Methodology*, 26, 39-78.

6.8.14 元分析(meta analysis)

这个术语指的是有关组织的累积证据的分析程序——既在单个研究内又横跨同一现象的多个研究——这种证据是由不同研究者在不同时间收集的。它是一个有价值的分析工具。Lipsey 和 Wilson 的下列讨论,摘录自他们的著作《应用元分析》(*Practical Meta-Analysis*),突出了这种工具的用途,以及它的优势和劣势。

关于元分析的推荐读物

Furnham, Adrian, & Hayward, Robert. (1997). A study and meta-analysis of lay attributions of cures for overcoming specific psychological problems. *Journal of Genetic Psychology*, 158, 315-331.

Glass, Gene V., McGaw, Barry, & Smith, Mary Lee. (1981). *Meta-analysis in social research.* Beverly Hills, CA: Sage.

Hunger, John E., Schmidt, Frank C., & Jackson, Gregg B. (1984). *Meta-analysis: Cumulating research findings across studies.* Beverly Hills, CA: Sage.

McLaughlin-Cheng, Elissa. (1998). Asperger syndrome and autism: A literature review and meta-analysis. *Focus on Autism and Other Developmental Disabilities*, 13,234-245.

McNamara, James F., Morales, Pamilla, Yeonbee, Kim, & McNamara, Maryanne. (1998). Conducting your first meta-analysis: An illustrated guide. *International Journal of Educational Reform*, 7, 380-397.

Ping, Xin Yan, & Jitendra, Asha K. (1999). The effects of instruction in solving mathematical word problems for students with learning problems: A meta-analysis. *Journal of Special Education*, 32, 207-225.

Ross, Steven. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing*, 15, 1-20.

Wolf, Fredric M. (1986). *Meta-analysis: Quantitative methods for research synthesis.* Beverly Hills, CA: Sage.

应用元分析

M.W.Lipsey & D.B.Wilson

元分析应用于哪种情形

元分析可以被理解为调查研究的一种形式,调查的是研究报告而不是人。编码者提出一个编码形式(调查协议),搜集研究报告的样本或者总体,且“面试”每篇研究论文。他认真阅读研究论文,对关于其特征和定量发现的适当信息进行编码。接着,通过使用传统统计方法的特定适当部分来调查和描述在所选研究中调查结果的模式,来分析结果数据。

元分析是各学科中对所选择的学术成果进行总结、整合和解释的众多方法中的一种,它有着一个重要的,但又有点局限的应用领域。首先,元分析只应用于经验研究论文,而不能用于总结性理论论文、常规的研究评论、政策建议等类似的论文。其次,元分析只应用于得出定量调查结果的研究论文,也就是,使用定量测量法和描述或者推论统计来总结数据结果的论文。这就排除了研究的定性形式,例如个案研究、民族志和“自然主义的”调查等。第三,元分析是对统计进行编码和分析的方法,这些统计所总结的研究发现是研究报告中结果的典型代表。如果我们所感兴趣的研究可以得到充分的数据,通常它将更适合于且利于直接使用传统的程序来分析它们,而不是利用元分析来总结统计值。

另外,由于元分析的关注点在于不同研究论文调查结果的聚集和比较,所以,那些调查结果可以有意义地比较是必要的。这意味着调查结果必须①概念上可比较,也就是,涉及相同构想和关系;②具有相似的统计形式。例如,如果可以通过不同治疗的互相比较来判断哪种是适当的,同时治疗结果具有相同的基本形式,即同时对对照组和控制组的回答者进行忧郁症的测量,那么一组忧郁症治疗效果的研究就可以进行元分析。在相同的元分析中包括显然不同的研究主题通常是不适当的,比如,忧郁症治疗研究和空间视觉的性别差异研究。这经常被称为元分析中的“苹果和橙子”问题——

企图总结或者整合并没有真正涉及相同构想和关系的研究。

同样地,若研究数据来自不同研究设计和以不同统计形式出现,即使其涉及相同主题,也不适合采用元分析。例如,使用治疗对控制组比较的忧郁症治疗的实验研究,通常不能与忧郁水平和获得服务水平相关的观察研究合并。虽然,两种研究以某种方式处理治疗和忧郁之间的关系,但是,研究设计的差异,构成结果的定量关系的本质,以及那些结果的意义是如此重要,以至于它们难以在同一个元分析中合并。当然,有人可能使用适当程序,分别对实验结果和相关结果进行元分析,并得出一些结论。

元分析的结果必须由对实践理由和概念理由的可比较的研究设计产生。元分析以效应规模的形式提出每个研究结果。效应规模是一个统计值,它从每个相关研究的结果中对临界的定量信息进行编码。不同类型的研究结果通常要求不同的效应规模统计值。例如,使用一个不同的效应规模统计值得出双变量相关的研究,是比对受试验组的因变量均值进行比较的研究更典型的元分析。类似地,报告一个单一主题样本前后均值差异的结果,使用的也是不同的效应规模统计值。此外,还存在其他更加专门化的例子。

假定有可比较的统计形式,对于什么研究结果在理论上可作元分析目的之比较的界定,通常只有支持者持固定的看法。与某一分析家绝对不同的结果可能看起来和另一个人相似。例如,Glass 的精神疗法效应的元分析(Smith & Glass,1977)被批评为合并来自明显不同治疗法的结果,如,认知行为、精神动力学、格式塔心理学等。Glass 声称他的兴趣在于精神疗法所有主要学派的全部效应和不同学派之间的比较,因此所有的不得不在元分析中提出。另一个分析家可能兴趣较窄,只用元分析研究结果,比方说,蛇恐怖症的脱敏治疗。但是,在任何一个例子中,分析家都需要界定其兴趣范围、包涵的基本原理以及排除在元分析之外的研究。有些分析家可能批评定义和基本原理,但只要他们是清楚的,每个批评家可以为他或者她自己判断它们是否有意义。

效应规模的关键概念

即使一组涉及相同主题并包括可比较研究设计的定量研究结果,仍会有一个重要的问题留给想把那些结果编码成一个可以进行有意义分析的数据库的人。一般情况下,这些研究对它们的关键变量将不使用相同的操作化(测量程序)。比方说,我们选择忧郁症治疗效应的比较组研究。一些研究可能使用 Beck 的忧郁总量作为结果变量,一些可能使用 Hamilton 的忧郁等级量表,一些可能使用临床医学家的忧郁等级量表,一些可能对这种构想拥有其他特殊的但有道理的测量。用这些完全不同的测量产生不同的有意义的数值,只与特定的操作和量表使用有关,他们的定量结果如何以允许它们统计合并和比较的方式进行编码?

答案涉及元分析的本质特征,确实,这个特征使元分析可能,并且提供环绕全部过程的中心。用于对元分析中定量研究结果的不同形式进行编码的不同效应规模统计是基于标准化的概念。效应规模统计提出研究结果的统计学标准化,以致作为结果的数值是可解释的,以一种稳定的方式越过所有的变量和有关的测量。当我们在检验和测量法谈到标准分时,标准化确实具有很多相同的意义。例如,我们可能把数学成绩

测验分转变成百分点,或者关于分值样本的标准差的标准化 Z 值,从而可以用一个不同的变量对它们进行有意义的比较,如,阅读成绩分。约翰可能数学在第 85 百分点但阅读只在第 60 百分点。

同样地,元分析中最常见的效应规模统计对兴趣测量分值样本分布的方差进行标准化。因此,关于 Beck 的忧郁总量治疗组和控制组之间均值差异可以按照标准差单位来表示,临床医学家的平均等级和所有这种其他忧郁定量测量也可以按照标准差单位来表示。在公制的标准差单位内,越过不同测量和操作化,人们也可以合并和比较结果。使用 Beck 的忧郁总量的研究可能显示出治疗组和控制组之间 0.30 的标准差差异,然而使用临床医学家的忧郁等级量表可能显示出 0.42 的标准差差异。假定各自样本的潜在总体是相同的,我们可以比较这些数字,在统计分析中用它们来计算均值、方差、相关和其他类似的结果,且通常把它们视为相同事情的有意义指标——在这一情况下,治疗组回答者体验忧郁的数量和控制组回答者体验忧郁的数量之间的差异就与忧郁估计总体的可变性相关。

因此,元分析的关键是定义一个能够以标准化形式提出一组研究论文定量结果的效应规模统计,这个标准化形式许可论文中有意义数字比较和分析。那儿有许多种可能性。效应规模的根本形式是把研究结果二分成统计上显著和统计上不显著。一个稍微更加细分的版本是每个统计显著检验的 p 值(如,$p = 0.03$,$p = 0.50$. Becker, 1994)。然而,这些不是非常好的效应规模统计。更合意的形式是关系的大小和方向,不仅仅是它的统计显著性。此外,它们被定义了,所以相对来说很少与其他问题分不清,诸如样本规模,它通常在显著检验的结果中明显地标出。

元分析家应该使用一个效应规模统计,它可以为其所考察的一组研究中所表述的特定的研究设计、定量结果的形式、变量和操作化提供适当的标准化。可用于一个情形或者另一个情形的效应规模统计是很多的,但实际上,只有很少几个可以得到广泛使用。最多经验结果属于几个普通分类的一种,这些普通分类发展和广泛承认明确的效应规模统计和相关的统计程序。

元分析的优点

之所以会考虑使用元分析来总结和分析研究论文的主体而不是传统研究评估技术,主要是基于四个元分析的主要优势。

首先,元分析程序施加有用的规律性于总结研究结果的过程。好的元分析作为一种凭自己本身能力进行的结构化的研究方法,因此要求每一个步骤都要证明和公开详细审查。它涉及一些具体标准,包括对界定研究发现之总体,对确认并检验合格研究的有组织的搜索策略,对研究特点与发现的正式编码,以及对支持所做结论的数据分析等等的标准。通过使研究总结过程清楚和系统,使用者可以评估作者的假设、程序、证据和结论,而不是对结论的有效性想当然的相信。

其次,元分析提出关键研究结果,在某种意义上,比依靠定性总结和统计显著的"投票计算"的传统检查程序,更具区分性和综合性。通过论文中每个相应统计关系大小和方向的编码,元分析效应规模构成一个对不同强度的研究结果敏感的变量。相

反,使用统计显著性来区分那些有效应的研究和无效应的研究的方法,潜在地存在着相当的误导。统计显著性反应了估计效应的大小和近似估计的样本误差,后者几乎完全是样本量的函数。因而,小样本研究可能因为低统计作用,发现有意义的影响或关系在统计上不具有显著性(Lipsey,1990;Schmidt,1992,1996)。

第三,元分析具有发现在总结研究的其他方法中模糊的效应或关系的能力。对结果的定性的、叙述性的总结,尽管能提供信息,但不能对它们自己结果中的研究与相关的差异之间的差别进行详细的审查。另一方面,元分析中典型研究特征系统的编码准许对研究结果与这种研究特征之间关系的分析作出精确的检查,如回答者特征、治疗本质、研究设计和测量程序等。此外,通过估计每个研究效应的量表和集中研究的那些估计(大规模的研究给予高的权重),较之个案研究,元分析提出用相对更多的统计影响力来综合效应估计。因此,元分析比较少系统分析的方法更可能发现与研究相一致的有意义的效应和关系,以及涉及研究差异的差异效应。

第四,元分析提供一个从正在检查的很多研究结果中处理信息的有组织的方法。当对从每个研究中获得的论文数量或者信息量要求比较低时,记笔记或者在索引条上编码不能有效地明了所有的细节。通过对比,元分析系统编码程序和记录结果信息的计算机化数据库的建构,几乎具有从每个研究中详述信息和覆盖很多研究的无限的能力。举个例子,由本书作者中的一人引导出的元分析带来了将近500个研究的每个研究中多于150个项目信息的数据库(Lipsey,1992)。然而,我们虽促进增加,但元分析不要求很多研究,且在某种情况下,它可以有效地应用于少到两个或者三个研究结果中。

元分析的缺点

元分析是有缺点的,这些缺点是来自某些方面严厉批评的主题(Sharpe,1997)。元分析的缺点之一在于它不过是付出一些努力和采用了一些技术。确切地说,很有几个研究指出元分析是劳动密集型的,且比传统的定性研究综述花费的时间更多。另外,元分析的许多方面要求专门的知识,尤其是适当效应量表的选择和计算及统计分析的应用。当然,这本书的主要目标是使专门的知识在实践层面对希望从事或者理解元分析感兴趣的人有用。

对元分析另一个关注涉及它的结构化的、有点机械的程序,这从另一角度可以视作它的优点。就一些应用(某些批评认为是对所有应用)来说,对研究数据元素和研究论文效应规模的相对客观编码,以及这些数据所能够进行的分析类型,可能对一些重要问题,比如研究的社会背景、理论的影响和推论、方法论的质量,以及设计、程序或者结论等更精细且复杂方面等等,失去了敏感性。利用以前使用的调查研究类推,元分析是一个总结研究结果的结构的、封闭式的问卷方法。一些调查应用要求一个更加开放的方法,比如,无结构访谈或者焦点团体等,以便于涉及某一主题的复杂性或者精细性。一些研究问题也要求一个比元分析能提供更定性的评估和总结。当然,为什么元分析和定性评论不能用于相同研究结果主体且得出所有结论,这在原则上是没有什么理由的。元分析的一个方法,Slavin(1986,1995)称之为最好的证据综合,尝试在相

同的研究评论中合并定性和定量的评论方法。

或许,对元分析最持久的批评与所包括的研究混合物有关(早先提到的苹果和橙子问题)。一些带着某种理由的批评认为,如果它们是将不能比较的研究结果集中起来的话,元分析提出的效应规模和其他总结统计是没有意义的。例如,为有关药物依赖者的镇静剂维持、社会技巧的性别差异和雇员士气的联合效果等研究结果的混合物建构效应规模分布,将没有意义。另一方面,很少人将反对从相同研究中实际复制结果所进行的元分析。然而,大多数关于这一点的批评显然未表现出如此极端的态度。当元分析家包括明显地既没有复制,又声称与更宽阔的主题相关的研究结果时,之间的灰色区变得有争议。正如更早提到的,Smith 和 Glass(1977)在他们初创元分析时以他们感兴趣的问题——精神疗法所具有的效力为基础,包括了一个大范围的精神疗法研究。然而,他们被研究者尖锐地批评,这些研究者发现不同治疗法与不同结果之间的巨大差异,并且觉得这种报告是对关于像行为治疗、精神动力治疗和格式塔治疗这些截然不同的方法和像害怕与焦虑、自尊、整体顺应、情感-肉体问题和学派功能这些不同的结果的平均效应的一种误导。

当然,当不同类型的研究结果被摊平聚合成一个基本平均规模效应时,问题就出现了。希望涉及广泛主题的元分析家日益把他们的任务处理为比较的一种,而不是集合。元分析提出了研究结果的明显不同的子范畴,它们可以各自准备,各自报告允许在它们中比较的效应规模分布和相关统计。此外,元分析的技术提升使同次性统计检验成为可能,此检验决定从不同研究中的分组效应量表,是否比单独来自抽样误差的预期显示了更多的偏差。这为是否研究显示了根本不同的结论,以至于没有理由假定它们是可比较的,提供了一个经验检验。不同的是,当代的元分析日益注意效应规模分布的方差而不是那些分布的均值。也就是,所关心的主要问题经常和识别研究结果差异的来源有关,而不是聚集结论成一个重要平均数。这个强调提供了一个更仔细处理研究结果的明显不同子群的方法,同时也少冒一些惹恼关心这些差异的批评家的危险。

相关且更麻烦的问题是,在同一个元分析中不同的方法质量的研究结果混合在一起。一些批评家认为研究综合只有以最高质量研究的结果为基础,且不应该被那些方法上有缺陷的研究得出的结论所降级。的确,元分析的一些方法为结论严格设置方法标准,比如,早先提到的最好证据综合方法(Slavin,1986)。使这点有争议的是,无论元分析采取什么样的方式,都存在有疑问的平衡和判断理由。一个困难是,除很少几个简单的标准以外,研究者之间相对地很少对方法性质的组成物达成一致意见。此外,很少研究领域提供所有评论者都认为是方法上没有缺点的,且使元分析有意义的研究。许多研究领域,尤其是那些涉及应用主题的领域,实质上提供的是不完美的研究和那些元分析家最感兴趣的但却是在没有代表性的情况下进行的、最接近教科书标准的研究。举个例子,在示范计划和大学诊所里,精神疗法方法上严格的研究比在常规精神健康实践里更可能(Weisz, Donenberg, Han, & Weiss, 1995; Weisz, Weiss, & Donenberg, 1992)。因此,我们对某些问题所具有的许多知识,存在于方法上有缺点的和潜在地易误解的研究中。元分析家必须决定,对于那些虽能判断说明但却有缺陷的

研究结果的结论能达到什么程度的赞同。因为他们知道,不严格的方法标准可能导致“垃圾研究的进进出出”,会令人嘲笑与责备,而严格的标准则又可能把关于一个主题的许多或者大多数可利用的证据排除在外。

关于这个问题已经形成两种方式:一种方式是严格控制方法标准并接受由此所带来的可资利用的相关研究发现的数量受到局限这个后果。在这种情况下,元分析家确信综合只以“最好”证据为基础,但它的结果可能只总结一个狭窄的研究领域且很少具有一般性。另一个方式是把研究中的方法变差作为一个经验问题对待,此经验问题将被作为元分析的一个部分进行考察(Greenland,1994)。在这种情况下,较不严格的方法标准是被加强的,但元分析家仔细编码方法的特征可能影响研究结果。于是,在统计分析的一个阶段,考察各种方法特征与研究结果相关的程度(如,在治疗研究中的概率对非概率分布)。如果一种可疑的方法实践与研究结果没有可论证的关系,那么相应的结果就包括在获得它们所包含证据的好处的最终的分析中。然而,如果可疑的实践研究显示出显著地不同于那些没有那个问题的结果,它们可能被排除在最后的结论之外,或者只使用统计判断来纠正它们偏差。

元分析的最近历史与当代应用

已发表的来自不同研究结果的定量综合的例子可以追溯到 1904 年,Karl Pearson 求得 5 个独立样本的伤寒发烧疫苗预防注射与死亡率之间相关的平均数(Cooper & Hedges,1994)。然而,正如在本章开头所提到的,元分析的现代时期开始于 Glass 关于精神疗法的著作(Glass,1976;Smith & Glass,1977;Smith,Glass,& Miller,1980),Schmidt 和 Hunter(1977)关于就业检验的有效系数,Rosenthal 和 Rubin(1978)关于人际期望效应。

大部分地受这些引人兴趣的应用的刺激,下一阶段的发展主要地集中在元分析的方法和统计的基础上。20 世纪 80 年代初期,以出版许多与书一般长度的概念、方法和元分析不同版本统计理论的出版物为标志(Glass,McGaw,& Smith,1981;Hedges & Olkin,1985;Hunter,Schmidt,& Jackson,1982;Light & Pillemer,1984;Rosenthal,1984;Wolf,1986)。这些书提供了实践和方法的指导,与初创著作的重要产生结合,引发元分析的应用及其评论的实质大发展。元分析在社会科学里迅速地传播开,尤其是教育学和心理学,且在健康科学领域以特别的活力流行起来,它实质上已经制度化成整合临床实验研究结果的首选方法(Chalmers et al. ,1987;Olkin,1992;Sacks et al. ,1987)。

其间,另一个更深奥的方法和统计著作的产生已经拓展并巩固了元分析的基础(如,Cook et al. ,1992;Cooper,1989;Hunter & Schmidt,1990;Rosenthal,1991)。这些成就的高峰是由罗素塞奇基金会(the Russell Sage Foundation)赞助的研究综合手册(*The Hand book of Research Synthesis*,Cooper & Hedges,1994),这是一本由 32 章组成的出版物,实质上包括了元分析的每一个方面,并且由这个领域的一批卓越的作者撰写。与这个背景相反,现代的任务是把最近的方法和统计著作整合和转变成实践指导,它能引导社会和行为科学领域经验研究结果的艺术的元分析。

资料来源:M. W. Lipsey and D. B. Wilson (2001). *Practical Meta-Analysis* (Thousand Oaks, CA: Sage) pp. 1-11,使用获得授权。

6.8.15 多元判别(函数)分析(multiple discriminant (fourdation) analysis)

判别函数方法已经由 Fisher 扩展到包括两组以上[参见 Fisher(1938)]。运用计算机,现在越过许多变量同时研究很多组已成可能的。

多元判别函数分析提供三种信息:

- 它决定事实上确定的组是否真的区别于选出的特征。
- 它指出根据什么因素可将各组最好地区别开来。
- 它指出一个个体是否类似于他或者她已指定的组中的其他个体。换句话说,它指出个体已经在多大程度上被理论地进行错误分类。

参考文献

Fisher, R. A. (1938). The Statistical utilization of multiple Measurements. Annals of Eugenics, 13, 376-386.

关于多元判别(函数)分析的推荐读物

Azari, N. P., Pettigrew, K. D., Pietrini, P., Murphy, D. C., Horwitz, B., & Schapiro, M. B. (1995). Sex differences in patterns of hemispheric cerebral metabolism: A multiple regression/discriminant analysis of positron emission tomographic data. *International Journal of Neuroscience*, 81, 1-20.

Azari, N. P., Pietrini, P., Horwitz, B., Pettigrew, K. D., Leonard, H. L., Rapoport, J. L., Schapiro, M. B., & Swedo, S. E. (1993). Individual differences in cerebral metabolic patterns during pharmacotherapy in obsessive-compulsive disorder: A multiple regression/discriminant analysis of positron emission tomographic data. *Biological Psychiatry*, 34, 798-809.

Chen, T., Bruininks, R. H., Lakin, K. C., & Hayden, M. (1993). Personal competencies and community participation in small community residential programs: A multiple discriminant analysis. *American fournal on Mental Retardation*, 98, 390-399.

DeRose, Daryn. (1992). Comparing classification models from multinomial logistic regression and multiple discriminant analysis. *Dissertation Abstracts International*, 52, 6696.

Klecka, William R. (1980). *Discriminant analysis.* Beverly Hills, CA: Sage.

Le Blanc, Louis A., & Rucks, Conway T. (1996). A multiple discriminant analysis of vessel accidents. *Accident Analysis and Prevention*, 28, 501-510.

Oleckno, William A., & Blacconiere, Michael J. (1990). Multiple discriminant analysis of smoking status and health-related attitudes and behaviors. *American Journal of Preventive Medicine*, 6, 323-329.

Rettig, Solomon. (1964). Multiple discriminant analysis: An illustration. *American Sociological Review*, 29, 398-402.

6.8.16 逆正函回归和对数比率模型(probit and logit models)

这些模型是非线性回归,作为一般多变量线性回归模型的选择使用,它们被用于估计一个二分因变量。这种二分因变量的常见例子有"成功"或者"失败","发生"或者"未发生","保持假释"或者"违反假释","在学"或者"辍学"。逆正函回归和对数比率方法是必要的,因为估计一个二分变量的一般最小二乘法回归是误导的。这违反了一般最小二乘法模型的一个关键假设——所有观察值都存在变差。

关于逆正函回归和对数比率模型的推荐读物

Aldrich, John H. , & Nelson, Forrest. (1984). *Linear probability, logit, and probit models.* Beverly Hills, CA: Sage.

Dey, Eric L. , & Astin, Alexander W. (1993). Statistical alternatives for studying college student retention: A comparative analysis. *Research in Higher Education*, 34, 569-581.

Gibbons, R. D. , & Lavigne, J. V. (1998). Emergence of childhood psychiatric disorders: A multivariate probit analysis. *Statistical Medicine*, 17, 2487-2499.

Hagle, Timothy M. , Mitchell, Glenn E. , II. (1992). Goodness-of-fit measures for probit and logit. *American Journal of Political Science*, 36, 762-784.

Rosenthal, James A. , & Rosenthal, Donald H. (1991). Logit and probit models: A juvenile justice program evaluation. *Social Work Research and Abstracts*, 27, 16-21.

6.8.17 最小空间分析(smallest space analysis)

这是一个绘图方法,是由 James C. Lingoes 和 Louis F. Guttman 开发的,用图解的术语来描述社会现象。社会对象根据它接近于选择的社会特征来描绘,结果可能用于一维、二维或者三维的空间。

关于最小空间分析的推荐读物

Cheng, Man Tsun. (1992). A smallest-space analysis of employment changes in Japan. *Sociological Perspectives*, 35, 593-627.

Foisy, Pierre. (1994). Age-related deficits in intentional memory for spatial location in small-scale space: A meta-analysis and methodological critique. *Canadian Journal on Aging*, 13,353-367.

Lingoes, James C. (1973). *The Guttman-Lingoes Non-metric program series.* New York: Academic Press.

Morrison, Paul, Burnard, Philip, & Hackett, Paul. (1991). A smallest space analysis of nurses' perceptions of their interpersonal skills. *Counselling Psychology Quarterly*, 4, 119-125.

Poreh, Amir, & Shye, Samuel. (1998). Examination of the global and local features of the Rey Osterrieth Complex Figure using faceted smallest space analysis. *The Clinical Neuropsychologist*, 12, 453-467.

6.8.18 随机过程(stochastic process)

在随机过程研究中,它一般关心随机变量的顺序,特别关于它们相依和限制的行为。随机过程至少部分由一些随机系统支配和由一个相应的数学模型表示。

在物理自然中,提供随机过程的例子是诸如细菌群体的总体增长。相似地,随机概率程序可能被用作人类流动、人口增长和移民的模型。

关于随机过程的推荐读物

Balakrishnan, J. D. (1994). Simple additivity of stochastic psychological processes: Tests and measures. *Psychometrika*, 59, 217-240.

Fararo, Thomas J. (1970). Stochastic processes. In Edgar F. Borgatta & George W. Bohrnstedt (Eds.), *Sociological methodology* 1969. San Francisco: Jossey-Bass.

McClelland, James L. (1991). Stochastic interactive processes and the effect of context on perception. *Cognitive Psychology*, 23, 1-44.

McMillan, N. J., Sacks, J., Welch, W. J., & Gao, F. (1999). Analysis of protein activity data by Gaussian stochastic process models. *Journal of Biopharmacy Statistics*, 9, 145-160.

Newell, K. M., Slobounov, S. M., Slobounova, E. S., & Molennar, P. C. (1997). Stochastic processes in postural center-of-pressure profiles. *Experimental Brain Research*, 113, 158-164.

Parzen, Emanuel. (1962). *Stochastic processes*. San Francisco: Holden-Day.

Rumsey, Deborah Jean. (1994). Nonresponse models for social network stochastic processes. *Dissertation Abstracts International*, 54, 4236B.

6.8.19 结构方程和结构方程模型 (structural equation and structural equation models)

结构方程把一个因变量和被认为对这个因变量有因果影响的各种结构的成分联系起来。可以发展各种结构方程模型。每个“因变量”必须明确地被认为是由系统中的一些联合变量所完全决定的。由测量变量完全决定是没有问题的,但必须介绍和其他决定变量没有关联的剩余变量。

举个例子,让我们假定一个阶层系统,在这个系统里报酬保证社会结构内各种职业位置的人的就业和动机。报酬可能包括声望、收入、闲暇和令人愉快的事物。声望被指定为因变量 X_3,且其他报酬如收入、闲暇和令人愉快的事物被指定为第二个因变量 X_4。现在假设我们假定一个职业规定的功能的重要性 X_1 和必须技巧 X_2 将分配声望、收入和闲暇以适当的报酬,并将引导积极的职责到实际的地位。

使用 X 来代表一个给定变量的标准值,可以以影响路径来表示关系,如:

$$\text{声望 } X_3 = p_{31}X_1 + p_{32}X_2 + p_{3a}X_a$$

$$\text{其他报酬 } X_4 = p_{41}X_1 + p_{42}X_2 + p_{4b}X_b$$

这些结构方程符合多元回归方程。在这些方程里,p 是一个路径系数(在标准多

元回归方程中路径系数等同于β系数)，且下标表示路径系数连接的变量。X_a和X_b属于外部于声望(X_3)和收入(X_4)的反映变量，也是可以影响因变量的测量误差。有时这些变量被称作剩余变量。这是一个完全递归模型，因为在X_1，X_2，X_3和X_4四个明确变量间画出了所有可能的单向箭头。

术语递归提到方程的系统(如上)，方程中任何一对变量间的相关可以沿着一般前置变量的路径写出。在路径图里，这是用单向箭头从每一个决定变量到每一个依赖于它的变量描绘出的。非递归系统包括变量瞬间的反馈行为。因此，这种系统无法绘出影响路径。

关于结构方程与结构方程模型的推荐读物

Bandalos, Deborah L. (1997). Assessing sources of error in structural equation models: The effects of sample size, reliability and model misspecification. *Structural Equation Modeling*, 4, 177-192.

De Jonge, Jan, & Schaufeli, Wilmar B. (1998). Job characteristics and employee well-being: A test of Warr's vitamin model in health care workers using structural equation modeling. *Journal of Organizational Behavior*, 19, 387-407.

Fan, Xitao, Thompson, Bruce, & Wang, Lin. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling*, 6, 56-83.

Goldberger, Arthur S., & Duncan, Otis Dudley (Eds.). (1975). *Structural equation models in the social sciences*. New York: Academic Press.

Hui, Chun, Law, Kenneth S., & Chen, Zhen Xiong. (1999). A structural equation model of the effects of negative affectivity, leader-member exchange, and perceived job mobility on in-role and extra-role performance: A Chinese case. *Organizational Behavior and Human Decision Processes*, 77, 3-21.

Seifert, T. L. (1997). Academic goals and emotions: Results of a structural equation model and a cluster analysis. *British Journal of Educational Psychology*, 67, 323-343.

7

ASSESSING SOCIAL VARIABLES: SCALES AND INDEXES

评估社会变量：量表与指数

准确说来，用以测量社会变量的量表与指数有数千种。即使已经有具备更高信度和效度的量表可用，社会科学家也会经常从中挑选以建构新的测量标准。这样做浪费时间，也浪费精力和金钱。另外，它会使我们发现，在研究的重复和累积过程中困难重重，甚至导致无法完成。本书所包含的指数的挑选是建立在效度、信度和效用等标准之上的。我们研究了社会测量中最常用的变量，并探讨了相关的测量方法。本书挑选了那些具有最高信度、效度和最大效用的量表。希望这本书能促进对这些量表更广泛的应用，或者促进研究者对更好的量表进行研究。

大致说来，在任何一项试图检验一个基本假设或者社会关系的研究设计中，有三组变量要素需要加以观察和测量。第一组是因变量，也就是我们想要观察和描述的结果。第二组是自变量，也就是被指定的原因性的要素。有时候自变量必须被分成或多或少是独立起作用的许多部分。第三组是干扰变量或者其他变量，这些变量必须加以控制以免它们使得我们希望通过实验来检验的关系变得模糊。

人们建构了许多量表来方便我们对于人类关系中的各种要素做定量描述。我们可以区分出三种社会测量的领域：

1）心理测量与社会心理学的量表：包括智力量表、人格测验与量表，以及态度测量与量表。这种量表的例子如明尼苏达多相人格测验（MMPI）和一些测量闲暇满意度、社区态度、成就取向和疏离感等的态度量表。

2）人口统计学量表：这些量表测量诸如社区、国家和民族等这样一些大的单位中社会行为的形成和结果。这种量表的例子包括社区评价量表、社区服务与居民政治活动量表以及社区团结指数等。

3）社会测量量表：这些量表被用于测量社会结构和过程。比如非正式友谊群体的社会测量和对于一些概念的测量，比如社会参与、社会距离和团体凝聚力等。其他量表则评估了婚姻协调程度和群体维度等。

我们需要谨记定性评估工具以及近来发展的定量评估工具在数量上的增长。如果你没有找到适合自己独特研究兴趣的量表，可以查阅《社会学评论年刊》(*Annual Review of Sociology*，章节:5.2.9)和《美国社会学评论》(*American Sociology Review*；章节:5.2.10)。也可以查阅其他期刊，比如《心理学季刊》(*Psychology Quarterly*)、《美国社会学杂志》(*American Journal of Sociology*)。

量表建构产生出了四种变量，反映了最早由实验主义心理学家 S. S. Stevens 揭示出来的不同水平的测量：定类变量，仅仅包括那些可以明显区分的类别，单独某个变量不能进行任何测量；定序变量，对于这类变量，一个等级排序中的位置可以区别出来，但是不同位置间的距离是没有意义的；定距变量，在一个连续体中任何两个相邻位置间具有相等的间距；定比变量，不仅具有相等的间隔，而且有一个绝对零点(意味着某一特性或者特征的缺失)。具体可参见本书表 6.1 对这些种类变量的总结。

理想状态是每一个变量都是定比变量，但是由于在测量某一特定生理物理现象的程序上可能存在一定的困难，所以现在所用的测量技术没有一种能满足定比量表的要求。定类变量不能排序也不能计算，它是一种基本的分类方案，因此这种量表通常被认为是在朝向社会变量的量化方向上的初级近似值。结果就是定序量表和定距量表最为常用。在关于定序量表或者定距量表是否为社会数据提供了最合适的模型的问题上存在着明显异议。一些作者认为现在所用的技术中几乎没有哪种能提供比定序量表更合适的数据，即使有也很少。另外一些人认为许多种量表可以被正确地看作是符合定距量表定义的。还有人则坚持，尽管目前所用的多数测量手段都没有超出定序量表，但是将适用于定距量表的统计方法用于这些量表应该是没有什么不利后果的。

其结果就是适用于定距量表的统计方法在社会数据的分析中继续得到广泛应用，不管这些数据是否符合对于定距量表的假设。特别适用于定序量表的统计方法也得到了越来越多的应用。本章第 4 节中所包含的统计工具是用于定序量表和定距量表的。

选择一个好的量表需要衡量许多标准。频率是选择一个量表时的有用标准。运用一个以前被频繁使用的量表，将会使累积起来的经验假设的研究成果变得更多，因为我们当前的研究是建立在以前累积起来的大量研究之上的。在文献中查阅标示该量表的关键词将会得到这一量表使用频率的粗略的近似值。

但是频率不是是否适用的惟一决定性因素，也不是最重要的。最重要的标准是效度。也就是说，这一量表是否测量到了它所要测量的东西。那么，为此有多少和什么样的证据能证明这一点，以及这一量表对于我们选择研究的问题恰当吗?

其他考虑因素包括信度(暗含在一个量表的效度中)、精度、简单及易操作性。近几年单维度研究受到了足够的重视。Guttman 技术使得研究者能够确定某一维度的建构量表。这在增加一个特定变量精度和可预测性方面应该是很重要的。但是我们必须要牢记两个条件：这一量表不管是在测量对于某一复杂对象的态度方面，还是在预测与该对象相关的行为方面，都不一定是最有效的。同样要记住，某一特定的量表对于某一群体的个体是单一维度的，对于别的群体则不是。

本书中这一部分所列的量表和评估工具是通过这样一些技术建构的：任意地或者

有判断性地排序法;目录分析技术;Thurstone 等距分析方法;Guttman 量表分析技术以及因素分析。不考虑在建构过程中所用的方法,研究者们所寻求的量表需要有最适合自己的问题,从而有最高的信度和效度,精确并且易于操作。当一个研究者选择某一量表来使用时,他必须了解随后将要用到的统计技术。通常,非参数统计用于定序量表,参数统计用于定距量表以及那些不太偏离随机性假设和正态分布的定序量表。

7.1 关于评估的资源概览

7.1.1 图 书

Bishop, Lloyd, & Lester, Paula E. (Eds.). (1993). *Instrumentation in education: An anthology*. New York: Garland.

Davis, Clive M. (1998). *Handbook of sexuality-related measures*. Thousand Oaks, CA: Sage.

Goldman, Bert Arthur, Mitchell, David F., & Egelson, Paula. (1997). *Directory of unpublished experimental measures*: 1990-1995 (Vol. 7). Washington, DC: American Psychological Association.

Johnson, Orval G. (1976). *Tests and measurements in child development: Handbook II* (2 vols.). San Francisco: Jossey-Bass.

Maddox, Taddy. (1997). *Tests: A comprehensive reference for assessments in psychology, education, and business* (4th ed.). Austin, TX: Pro-Ed.

Murphy, Linda, Plake, Barbara S., & lmpara, James C. (Ed.). (1995). *Tests in print V: An index to tests, test reviews, and the literature on specific tests.* Lincoln: Buros Institute of Mental Measurements, University of Nebraska at Lincoln.

Robinson, John P., Shaver, Phillip R., & Wrightsman, Lawrence S. (Eds.). (1991). *Measures of personality and social psychological attitudes.* New York: Academic Press.

Robinson, John P., Shaver, Phillip R., & Wrightsman, Lawrence S. (Eds.). (1998). *Measures of political attitudes.* New York: Academic Press.

Touliatos, John, Perlmutter, Barry F., & Straus, Murray A. (Eds.). (2001). *Handbook of family measurement techniques.* Thousand Oaks, CA: Sage.

7.1.2 出版社及测试供应者

American Guidance Service
4201 Woodland Road
Circie Pines, MN 55014-1796
800-328-2560
www. agsnet. com

Association of Test Publishers
1201 Pennsylvania Avenue
Suite 300

Washington, DC 20004
202-857-8444
www. testpublishers. org

Consulting Psychologists Press, Inc.
Davies-Black Publishing
3803 East Bayshore Road
EO. Box 10096
Palo Alto, CA 94303
650-969-8901
www. cpp-db. com

ETS Test Collection
Educational Testing Service
Princeton, NJ 08541
609-734-5686
www. ets. org

Health and Psychosocial Instruments (HAPI)
Behavior Measurement Database Services
P. O. Box 110287
Pittsburgh, PA 15232-0787
412-687-6850

MultiHealth Systems
P. O. Box 950
North Tonawanda, NY 14120-0950
800-456-3003

NCS Pearson
Operations Center
Atrium E 80, Route 4
Paramus, NJ 07652-2622
201-291-0233
www. ncs. com

Psychological Assessment Resources, Inc.
16204 N. Florida Ave.

Lutz, FL 33549
813-968-3003
www. parinc. com

The Psychological Corporation
19500 Bulverde
San Antonio, TX 78259
800-872-1726
www. psychcorp. com

PsycINFO Database
American Psychological Association
750 First Street
Washington, DC 20002
202-336-5500
www. apa. org/psycin fo/

Riverside Publishing
425 Spring Lake Drive
Itasca, IL 60143-9921
800-323-9540

7.2 评估社会地位的量表

社会阶层或者社会地位是社会研究中最重要的变量之一。一个人的社会经济地位影响了他的受教育机会、收入、职业、婚姻、健康以及朋友圈,甚至影响了预期的寿命。在一个像美国这样的多元的、平等的和流动的社会中测量这样一个变量是很困难的。因此一些研究者试图区分社会阶层并测量与之相关的变量。人们发现职业是社会地位的最好的单一预测指数,而总体而言,职业声望是高度稳定的,在职业与社会地位的关系中许多因素同时起着作用。我们知道个人收入和教育程度均与职业等级相关。教育是进入许多职业的基础,而对于大多数人而言,收入的多少又来自职业。房屋类型与居住地区构成了其他高度相关的因素。

这里给出了六种量表以供研究者在研究中参考或使用。各种量表的长度与其中包含的因素的数量是不同的。这些量表是:

- Duncan 的社会经济指数
- Siegel 的声望评分

- Treiman 的标准国际职业评价量表
- Nam-Powers 的社会经济地位(status)评分
- Hollingshead 的社会位置(position)指数
- 教育量表

7.2.1 Duncan 的社会经济指数

被测变量 这一社会经济指数与一些基本特征相关,比如职业声望、教育和收入。

描述 这一测量方法使研究者发展了两个客观事实:①在 1950 年人口调查的详细分类中,将 North-Hatt(全国民意研究中心,NORC)职业声望评价中的职业从 90 种扩展到 446 种;②获取一种职业声望指数以测量"NORC 声望评价与人口的社会经济特征"之间的关系。就其构成变量而言,这一指数具有表面的效度。考虑到 NORC 职业声望排名,它又具有充分的预测效率。

出处 初始量表:A. J. Reiss , O. D. Duncan, Paul K. Hatt 和 C. C. North 合著的《职业与社会地位》(*Occupations and Social Status*. Glencoe,IL:Free Press,1961)。修正的量表:R. M. Hauser, D. L. Featherman 合著的《分层过程:趋势与分析》(*The Process of Stratification*:*Trends and Analysis*. New York:Academic Press,1977)。修正的量表:D. L. Featherman, G. Stevens 通过采用最新的对劳动力收入和教育程度的测量而发展了 Duncan 的社会经济指数,在考虑到男性及全体劳动力的贡献的同时为职业声望的测量提供了一个更好的近似方法。这一修正以《一种修正的关于职业地位的社会经济指数——在成就的性别差异分析中的应用》(*A revised Socioeconomic Index of Occupational Status*:*Application in Analysis of Sex Differences in Attainment*)为题发表在《社会结构与行为——纪念 W. H. Sewell 文集》(*Social Structure and Behavior*: *Essays in Honor of William Hamilton Sewell*. New York:Academic Press)一书的第 7 章。

信度 一个人的职业可以根据美国职业与工业调查指数,以及由美国劳工部编写的《职业名称词典》(*Dictionary of Occupational Titles*)转译成职业代码,而信度问题的产生就是对这一转译的回应性描述。

效度 声望变量与每一个预测指数都高度相关,与教育相关度为 $r=0.84$,与收入为 $r=0.85$,而三个变量间的复合相关指数 $R_{1(23)}$ 为 0.91。总体职业评价能够长时间在不同社会系统中保持高度稳定(从 1947 到 1963,$r=0.99$)。(Hodge,Selgel & Rossi:1964)

Hauser 和 Featherman(1977)在分析 Duncan 的社会经济指数研究父亲的教育、职业和儿子的教育、第一份工作及当前的工作之间关系时提出了许多确认系数。人们发现在 Duncan 的社会经济指数与 Hollingshead 的社会地位双因素指数间存在中度相关($r=0.74$),然而这两种关于社会地位的主要测量方法出现这种差异是很自然的。

效用 所需要的基本数据是研究对象对他们职业的描述。这些数据被研究者转译成职业代码,职业名称则转换成已有的 Duncan 的社会经济指数(如果需要则转换成修正的 NORC 指数)。在职业编码中有许多很细致的地方,特别是关于蓝领职业。研究者应该留出足够的训练与交叉检查的时间。每一个打算得到详细的职业编码的

人都应该向 Bureau of Labor Statistics Standard Occupational Classification System 咨询。

参考文献

Hauser, Robert M., & Featherman, David L. (1977). *The process of stratification: Trends and Analysis.* New York: Academic Press.

Hodge, Robert W., Siegel, Paul M., & Rossi, Peter R. (1964). Occupational prestige in the United States, 1925-1963. *American Journal of Sociology*, 70, 286-302.

Hollingshead, August B. (1971). Commentary on "The indiscriminate state of social class measurement." *Social Forces*, 49, 563-567.

研究应用

Airsman, Linda A., & Sharda, Barn Dev. (1993). A comparative study of the occupational attainment processes of white men and women in the United States: The effects of having ever married, spousal education, children and having ever divorced. *Journal of Comparative Family Studies*, 24, 171-187.

Caston, Richard J. (1989). Dimensions of occupational inequality and Duncan's Socioeconomic Index. *Sociological Forum*, 4, 329-348.

Davey, Lynn F. (1993, March). *Developmental implications of shared and divergent perceptions in the parent-adolescence relationship.* Paper presented at the biennial meeting of the Society for Research in Child Development, New Orleans, LA.

Dworkin, Rosalind J. (1981). Prestige ranking of the housewife occupation. *Sex Roles*, 7(1), 59-63.

Ell, Kathleen, Larson, David, Finch, Wilbur, Sattler, Fred, et al. (1989). Mental health services among ambulatory patients with human immunodeficiency syndrome infections. *Journal of Health and Social Policy*, 1(1), 3-17.

Ganzeboom, Harry B. G., De Graaf, Paul M., & Treiman, Donald J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, 21(1), 1-56.

Giudicatti, V., & Stening, B. W. (1980). Socioeconomic background and children's cognitive abilities in relation to television advertisements. *Journal of Psychology*, 106(2), 153-155.

Gottfried, Allen W. (1985). Measures of socioeconomic status in child development research: Data and recommendations. *Merrill-Palmer Quarterly*, 31 (1), 85-92.

Graham, Sandra, & Long, Anna. (1986). Race, class, and the attributional process. *Journal of Educational Psychology*, 78(1), 4-13.

Hanson, Sandra L. (1982). The effects of rural residence on the socio-economic attainment process of married females. *Rural Sociology*, 47(1), 91-113.

Hirschman, Charles, & Kraly, Ellen Percy. (1990). Racial and ethnic inequality in the United States, 1940 and 1950: The impact of geographic location and human capital. *International Migration Review*, 24(1), 4-33.

Hodge, Robert W. (1981). The measurement of occupational status. *Social Science Research*, 10, 396-415.

Lin, Yangjing, & Vogt, W. Paul. (1996). Occupational outcomes for students earning two-year college degrees: Income, status, and equity. *Journal of Higher Education*, 67(4), 446-475.

Markides, Kyriakos S., Martin, Harry W., & Sizemore, Mark. (1980). Psychological distress among elderly Mexican Americans and Anglos. *Ethnicity*, 7, 298-309.

Mueller, Charles W., & Parcel, Toby L. (1981). Measures of socioeconomic status: Alternatives and recommendations. *Child Development*, 52(1), 13-30.

Mutchler, Jan E., & Poston, Dudley L. (1983). Do females necessarily have the same occupational status scores as males? A conceptual and empirical examination of the Duncan Socioeconomic Index and Nam-Powers Occupational Status Scores. *Social Science Research*, 12,353-362.

Stevens, Gillian, & Featherman, David L. (1981). A revised socioeconomic index of occupational status. *Social Science Research*, 10, 364-395.

7.2.2 Siegel(NORC)的声望评分

被测变量 一个测量分层系统的声望评分。

描述 对于职业声望的North-Hatt NORC研究出现于1947年,随后就在声望研究中得到了广泛应用。通过对美国成年人口的一次全国性抽样得到了90种职业的声望评分。1963年在国家科学基金会的资助下,NORC进行了同样的一次研究,目的是为一个更具有代表性的职业样本提供确切的声望评分,并且揭示出一些产生这种声望评分的职业特征。同在1947年那次研究中一样,职业评价的获得是通过让被调查者回答这一问题——"对于下列每一项工作,请挑选出最能反映你关于这种工作的地位的观点的选项"。被调查者因此判断一种职业具有"优"、"良"、"一般"、"低于一般"或者"差"的地位(还有一个"不知道"选项)。声望评价的一个标准是那些评价这一职业的被调查者回答"优"或者"良"的比例。

另一种测量方法出自一个职业排位的总体。这一总体是通过以任意数值来衡量不同的回答而形成的。确定了一个合适的排位后,对所有给职业排位的被访者给出的不同答案随机赋予的数值的平均值就产生了NORC声望评分。尽管由于分配给五种可能排位的数值衡量存在任意性,但后一种测量方法得到了广泛应用。

1963年的NORC职业声望排名仅限于90种职业,而在Duncan的社会经济指数中则有超过500种职业评分。到了1970年随着Hodge,Siegel和Rossi确定了超过400种声望评分,这一局限性消失了。最近,威斯康星的研究者将1964—1965年的NORC声望评分(出自Siegel发表在1971年关于1960年社会调查取得的详细职业名目的报告中)转换成了1970年的分类系统。

出处 《美国职业结构中的声望》(*Prestige in the American Occupational Structure*),作者:P. M. Siegel,(博士论文,芝加哥大学社会学系,1971年)

信度 研究人员发现被访者的社会地位和被访者对于各种职业的熟悉程度对评分存在着影响。Hatt发现8种职业或职业群体具有内在同质性:公务员、专业人员、商业、娱乐业、农业、手工业、军事组织和服务业。

效度 据Siegel报告(1971),Duncan的社会经济指数(1960)评分与NORC声望评分之间的相关系数为0.87。而NORC声望评分与Treiman的标准国际职业声望评分的相关系数为0.95。Duncan的社会经济指数似乎低估了职员、农民和某些蓝领工人(如操作工、木匠等)的排名而高估了企业家、新闻从业人员、清洁工的排名。与

Duncan 的排名相比,NORC 评分对这些职业提供了不同的排名,因此在研究这些职业的时候研究者倾向于使用 NORC 的体系。

由于 Duncan 声望评分是在同时权重了收入和受教育程度的基础上产生的,所以 Siegel 的声望评分经常被认为是关于职业的相对声望的估计,而这与实际的声望评分之间的近似度非常不理想。人们已经指出了一些明显的差异,而最重要的发现是,与 Duncan 的社会经济指数评分相比,Siegel 的声望评分在研究以职业流动为基础的地位维度中,它是一个缺乏效度的评分。其实,声望测量对于研究非社会经济性职业维度可能更为有效(Houser & Featherman,1977 年,pp. 37,50)。

效用 最好的做法是用别的方法给职业编码,并调查所得结果中的差异。如果研究者只希望停留在美国数据的单一职业地位评分上,那么 Duncan 指数具有两个优点:它应用最广泛并且可用于比较分析,它比声望评分更能抓住教育和收入之间的交叉变量。如果需要得到关于个人的社会经济地位的其他信息,那么 NORC 声望排名就可以避免收入与教育之间的人为夸大的相关性(Treiman,1977,pp. 311-312)。

参考文献

Hauser, Robert M., & Featherman, David L. (1977). *The process of stratification: Trends and analysis.* New York: Academic Press.

Hodge, Robert W., Siegel, Paul M., & Rossi, Peter H. (1964). Occupational prestige in the United States, 1925-1963. *American Journal of Sociology*, 70, 286-302.

Robinson, John P., Athanasiou, Robert, & Head, Kendra B. (1969). *Measures of occupational attitudes and occupational characteristics.* Ann Arbor: University of Michigan, Institute for Social Research.

Siegel, P. M. (1971). *Prestige in the American occupational structure.* PhD dissertation, University of Chicago Department of Sociology.

Treiman, Donald J. (1977). *Occupational prestige in comparative perspective.* New York: Academic Press.

研究应用

Bose, Christine E., & Rossi, Peter H. (1983). Gender and jobs: Prestige standings of occupations as affected by gender. *American Sociological Review*, 48, 316-330.

Featherman, David L., & Hauser, Robert M. (1976). Prestige or socioeconomic scales in the study of occupational achievement? *Sociological Methods and Research*, 4,403-422.

Glenn, Norval D. (1975). The contribution of white collars to occupational settings. *Sociological Quarterly*, 16(2), 184-189.

Kamieniecki, Sheldon, & O'Brien, Robert. (1984). Are social class measures interchangeable? *Political Behavior*, 6(1), 41-59.

Moreland, Carol P. (1978). *Sex and occupational attainment: A comparison of black males and females.* Paper presented at the meeting of the Mid-South Sociological Association.

Nilson, Linda Burzotta. (1978). The social standing of a housewife. *Journal of Marriage and the Family*, 40, 541-547.

Rytina, Steven L. (n. d.). *Scaling the intergenerational continuity of occupation: Is occupational inheritance ascriptive after all?* Unpublished manuscript, Department of Sociology, Harvard University.

Spaeth, Joe L. (1978). Measures of occupational status in a special population. *Social Science Research*, 7(1), 48-60.

Sullivan, Teresa A. (1983, April). *The occupational prestige of women immigrants: A comparison of Cubans and Mexicans.* Paper prepared for presentation at the annual meeting of the Population Association of America, Pittsburgh, PA.

Villemez, Wayne J., & Silver, Burton B. (1977). Occupational situs as horizontal social position: A reconsideration. *Sociology and Social Research*, 61,320-336

7.2.3 Treiman 的标准国际职业声望量表

被测变量 一种标准化的声望测量,可用于任何一个国家的职业编码,并做跨国比较。

描述 这一量表包括 509 种职业,288 个初级群体,84 个次级群体和 11 个主要类别的声望评分。这一量表覆盖 92 个点,从国家领导人(分数为 90)到税务员(分数为 -2)。而这 509 种职业的平均得分为 43.3。

出处 《比较视角中的职业声望》(*Occupational Prestige in Comparative Perspective*),作者:D. J. Treiman,(New York:Academic Press,1977)。

信度 这一量表已经显示出了高度的信度。7 个国家间平均相关系数为 0.97。

效度 标准量表与 55 个国家的声望等级有极高的相关性。根据 55 个国家的纯声望数据计算所得的国家间相关系数与标准量表间的平均相关系数为 0.97。在这 55 个国家中,只有 7 个国家在标准声望上表现出了小于 0.87 的相关性,对于美国和英国,相关性超过了 0.96。

效用 由于在所有社会中的声望评价基本相似,Donald Treiman 得以第一个创建能够有效地给所有国家中的职业分配声望分数的声望指数。

研究应用

Ganzeboom, Harry B. G., & Treiman, Donald J. (1996). Internationally comparable measures of occupational status for tile 1988 International Standard Classification of Occupations. *Social Science Research*, 25(3), 201-239.

Sharda, Bam Dev, & Elder, Joseph W. (1975, August). *Industrialization and the processes of stratification in rural societies: A comparison of rural India and rural United States.* Paper presented at the annual meeting of the American Sociological Association.

Spaeth, Joe L. (1978). Measures of occupational status in a special population. *Social Science Research*, 7(1), 48-60.

Treiman, Donald J. (1975). Problems of concept and measurement in the comparative study of occupational mobility. *Social Science Research*, 4(3), 183-230.

Treiman, Donald J. (1976). A standard occupational prestige scale for use with historical data *Journal of Interdisciplinary History*, 7(2), 283-304.

Wolf, Chnstof. (1997). The ISCO-88 International Standard Classification of Occupations in cross-national survey research. *Bulletin de Methodologie Sociologique*, 54, 23-40.

Yogev, Abraham. (1980). Validating the applicability of the Standard International Occupational Prestige Scale: Illustrations from Costa Rica and Israel. *International Review of Modern Sociology*, 10(1), 15-35.

7.2.4 Nam-Powers 社会经济地位评分

被测变量 这是一个能够测量职业、教育和家庭收入等组成项的平均值的多项目测量方法。同时它还可以得到对于地位稳定性的测量结果。

描述 20 世纪 60 年代,Charles B Nam 和 Mary G Powers 与一些在统计局工作的同事,以 1960 年关于收入和教育的数据为基础设计了关于职业的社会经济地位指数(未使用声望评价)。调查局的这些人认为不仅可以通过本质特征给职业分类,而且还可以通过设计一种将教育及收入的独立排位同职业排位结合在一起的关于社会经济地位的多项目指数,这样同质性可以更好地得到保证。

用于计算这些指数的程序与 Duncan 用过的有些相似,但是有下列不同之处:①采用的是教育和收入的中位数,而不是特定的教育和收入程度的比例;②Duncan 间接的以年龄分类对指数进行了标准化;③Duncan 采用的是 1947NORC 声望排位。然而据此前的报道,统计指数与 Duncan 的指数之间的相似程度被 Pearson 系数(0.97)证明了。在对 1970 年的社会调查设计方案时,统计局决定放弃使用任何措施来测试这些社会经济指数。

Nam 和 Powers 认识到了纯声望研究方法(比如:Siegel 的 NORC 声望指数)和社会经济地位(SES)的职业声望定位之间的差异。他们认为不考虑声望,对社会经济地位进行直接测量,这种第三方方案应该值得更好的考虑。这一方法在 August B. Hollingshead 和 Frederick C. Redlich(1958)的著作里,以及 Peter M. Blau 和 Otis Dudley Duncan(1967)的著作中也被采用过。这种方法是由于研究者出于这样的考虑:需要对地位或生活机会进行测量,或是对客观地位状况进行测量,所以任何评价职业的标准都会将研究者的注意力引向职业排位的纯社会经济指数上去。在上个世纪美国统计局的工作中,第三种纯粹社会经济的方法已经可以很明显地看出来它的优势。

Nam 和 Terrie 在 1980 年调查数据的基础上已经开始继续创建新的 Nam-Powers 职业地位指数。关于它们的纯粹社会经济多项目测量的基本原理,在 Nam 和 Powers1983 年的著作中可以找到。

出处 最初是基于 1960 年调查数据的著作,是美国统计局的研究者所著《方法论与社会经济地位得分》(*Methodology and Scores of Socioeconomic Status*,华盛顿特区:政府出版办公室,1963),以及美国统计局的《美国人口统计,1960,社会经济地位,最后报告 PC(2)-5C》(*U. S. Census of Population*, 1960, *Socioeconomic Status*, *Final Report PC*(2)-*5C*,华盛顿特区:政府出版办公室,1967)。在 1970 年的调查数据基础上,调查报告合并了男性、女性、全年全职女工以及两性的 SES 指数。这些指数可以在 Charles B. Nam,John LaRocque, Mary Powers 和 Joan Holmbergde 的文章《职业地位得分:稳定

与变化》(Occupational Status Scores: Stability and Change,文章发表于:*Proceedings of the American Statistical Association, Social Statistics Section*,pp. 570-575,1975)中找到。

信度 当 Nam 和 Powers 将 1950 年和 1960 年男性的详细职业列表整体对比之后,他们算得两套指数的相关系数为 0.96。采用了 126 种职业的 1950—1960 年调查数据的相关系数为 0.95。在 1960 年和 1970 年的数据中,对男性的 126 种职业的相关系数计算得到 0.97,表明在地位指数中能够发现很高的稳定性。1950 年和 1970 年调查数据中男性的指数之间的相关系数为 0.91。对于所有女性而言,1960—1970 年的系数为 0.85,虽然这样的结果还是很高的,但是比起男性来还是低很多。在 1980 年,人们发现了男性指数和女性指数之间的更具相似之处。这样,一套整合了男性和女性指数的得分体系就产生了。

效度 与 Duncan 和 Siegel 的声望测量方法相比,这种方法具有很高的相关性。尽管社会经济地位和声望这两个维度是高度相关的,仍然有这样一种趋势,即随着时间的变化职业地位比职业声望的变化更显著(Pavalko,1971,p. 132,p. 140)。

效用 女性现在日益融入社会劳动力中,将职业标示为"男性的"或"女性的"变得越来越不合适。

参考文献

Blau, Peter M., & Duncan, Otis Dudley. (1967). *The American occupational structure.* New York: John Wiley.

Hollingshead, August B., & Redlich, Frederick C. (1958). *Social class and mental illness: A community study.* New York: John Wiley.

Nam, Charles B., & Powers, Mary G. (1983). *The socioeconomic approach to status measurement.* Houston: Cap & Gown.

Pavalko, Ronald M. (1971). *Sociology of occupations and professions.* Itasca, IL: Peacock.

研究应用

Broman, Sarah H., Nichols, Paul L., & Kennedy, Wallace A. (1975). *Preschool IQ: Prenatal and early developmental correlates.* New York: John Wiley.

Chiricos, Theodore G., & Waldo, Gordon P. (1975). Socioeconomic status and criminal sentencing: An empirical assessment of a conflict proposition. *American Sociological Review*, 40, 753-772.

Myrianthopoulos, N. C., & French, K. S. (1968). An application of the U. S. Bureau of the Census socioeconomic index to a large, diversified patient population. *Social Science and Medicine*, 2, 283-299.

Nam, Charles B. (1968). Changes in the relative status level of workers in the United States, 1950-1960. *Social Forces*, 47(December), 167-170.

Nam, Charles B., & Powers, Mary G. (1965). Variations in socioeconomic structure by race, residence, and the life cycle. *American Sociological Review*, 30(February), 97-103.

Nam, Charles B., Powers, Mary G., & Glick, Paul C. (1964). *Socioeconomic characteristics of the population*: 1960 (Current Population Reports, series P-23, no. 12). Washington, DC: Government Printing Office.

Powers, Mary G., & Holmberg, Joan J. (1978). Occupational status scores: Changes introduced by the inclusion of women. *Demography*, 15, 183-204.

7.2.5 Hollingshead 的社会位置指数

被测变量 个人在地位结构中所处的位置。

描述 这一指数的双因素和三因素形式都得到了广泛应用。双因素指数由一个职业量表和一个教育量表构成。三因素指数还包含一个与居住相关的量表。由于与居住相关的量表是建立在对 New Haven 和 Connecticut 的社会学分析的基础上的，对于许多社区是不适合的，除非居住区域被绘制为一个六维量表。双因素指数只需要关于职业和教育的知识。

职业量表是一个七个点的量表，是关于将职业分类为许多社会经济群体的 Edwards 系统的修正。Edwards 系统并不区分职业的种类或者商业大小同商业的经济力量之间的差别。而 Hollingshead 社会地位指数则将职业者分成不同的群体并将商业以规模和价值排列。

教育量表同样分为七个方面。在双因素指数中，职业被赋予的权重为 7，而教育被赋予的权重为 4。如果要计算一个上过高中加一年贸易大学的 Kroger 蔬菜店的老板的分数，步骤如下：

因素	量度分数	×	因素权重	=	局部分数
职业	3		7		21
教育	3		4		12
			社会位置指数分数		33

出处 August B. Hollingshead 所著的《社会地位的双因素指数》(*Two Factor Index of Social Position*. 公开发表于 1957 年)。Yale Station, New Haven, CT. August B. Hollingshead 和 Frederick C. Redlich 所著的《社会阶级与心理疾病：对一社区的研究》(*Social Class and Mental Iuness*：*A Community Study*. New York：John Wiley，1958，pp. 387-397)。

关于 Hollingshead 对双因素量表的背景和基本原理的介绍见他所写的《对于社会阶级测量中不加选择的状态的说明》(*Commentary on The Indiscriminate State of Social Class Measurement*)，发表于 1971 年的 *Social Forces*，49 期，pp. 563-567。

社会位置指数的信度与效度：人们发现在 Hollingshead 和 Redlich 的测量方法与由 Ellis，Lane 和 Olesen 创建的阶层位置指数(The Index of Class Position，1963)之间具有高度的相关。

在 Guttman 意义上职业和教育的量表指数的不同组合是可再现的，因为在“教育—职业”组合中没有重合之处。如果一个人的教育和职业是已知的，我们就可以计算他或者她的分数；如果我们知道一个人的分数，我们就可以算出职业和教育等级。

Hollingshead 和 Redlich 报告所研究的阶层、教育同职业之间的相关度为 $R_{1(23)} = 0.906$。所研究的阶层与住所、教育和职业的相关度为 $R_{1(234)} = 0.942$。Hollingshead 和其他人通过使用超过 100 个的变量对于计分的信度及该指数的效度进行了广泛的研究。

Slomczynski, Miller 和 Kohn(1981)宣布关于职业地位的 Hollingshead 指数在用于研究美国的时候，通过纵向测量模型得到了确证，该模型显示 Hollingshead 指数同 Treiman 的国际声望量表、Hodge-Siegel 指数以及 Duncan 社会经济指数在职业地位方面是同样强有力的指数。

效用 由于完善的生态地图不存在，获取与居住相关的信息存在困难，因此只有社会地位指数的双因素形式得到广泛应用。这种方法只需要职业和教育，这些数据是相对容易获取的，那么量表指数就可以快速计算得出从而确立个人的社会地位。

参考文献

Ellis, R., Lane, W., & Olesen, V. (1963). The Index of Class Position: An improved intercommunity measure of stratification. *American Sociological Review*, 28, 271-277.

Slomczynski, Kazimierz M, Miller, Joanne, & Kohn, Melvin L. (1981). Stratification, work, and values: A Polish-United States Comparison. *American Sociological Review*, 46, 720-744.

下面的量表(双因素社会地位指数的职业量表)是得到了 Hollingshead 和 Redlich 的允许，复印自《社会阶级与心理疾病——对一个社区的研究》，7.2.6 部分的教育量表的有关内容也同样来自这一作品。

职业量表

1. 大公司的高级主管人员，经营者和主要专业人员
   A. 高级主管人员(Dun 和 Bradstreet 认为资产超过 500 000 美元的公司)
      银行
         行长
         副行长
         助理副行长
      商业机构
         副总裁
         助理副总裁
         执行秘书
         研究主管
         财务人员
   B. 经营者(Dun 和 Bradstreet 认为资产超过 100 000 美元的)
      经纪人
         承包人
         牛奶场场主
      农场主
      木材经销商
   C. 主要专业人员
      会计(加拿大太平洋航空)
      保险精算师
      农学家
      法官(高等法院)
      律师
      冶金家

审计师
建筑师
艺术家、肖像画家
天文学家
细菌学家
化学工程师
化学家
牧师(受专业培训的)
牙科医生
经济学家
工程师(大学毕业)
林务员
地质学家
军官(准尉、少校或者更高)
国家、联邦、州或是地方行政部门的官员(例如:城市建设局长、国家税务局长)
内科医师
物理学家
心理学家
交响乐团指挥
教师
兽医

2. 企业经理人,中等企业的经营者,较少收入的专业人员

A. 大公司的经理人(年薪50万美元的)

广告总监
部门主管
经纪人
购物向导
地区负责人
执行助理
国际财团出口部门经理、政府官员和普通职员例如:
国家税务人员
制造商代理人
办公室主管
人事主管
警察局长、县治安官
邮局女局长
制片人
销售专家
销售主管、国营公司
商店经理

B. 中等规模公司的经理人(年薪35 000~100 000美元)

广告公司
服装商店
承包人(商)
快递公司
农场主
水果批发市场
家具商行
珠宝商
家禽生意
不动产经纪人
地毯商人
百货商店
剧场

C. 较少收入的专业人员

会计师(不是CPA)
手足病医生
脊椎指压治疗者
校员
团体理事标
工程师(不是大学毕业)
军人:服役军官、陆军上尉、上校
音乐家(交响乐、管弦乐)
护士
光学仪器商
验光师
药剂师

财经作家
健康指导师
劳工关系顾问
图书管理员
公共健康官员
大学研究助教(全职)
社会工作者

3. 行政职员,小商行所有者,二流专业人士

A. 行政职员

广告代理商
主要职员
信用卡管理员
保险代理
部门主管
铁路乘务员
私人秘书
商品代理商
销售代表
地区代表,联邦、州、地方政府办公室
地区代表,大商业、工业公司
服务行业经理
商店经理
连锁商店经理
运输行业经理

B. 小商行主

艺术画廊
汽车零配件
遮阳篷
面包店
化妆品店
造船厂
保险经纪人
汽车经销商
家养牲畜经销商
香烟机器
清洁商店
服装
煤生意
缔约买卖
疗养院
装修工
宠物商店
脱水食物
雕刻艺术品商店
饲养行业
当地的金融公司
消防员
出售廉价小商品的杂货店长
种花人
家具
修车厂
加油站
玻璃器具类
普通食品杂货店
旅馆保安
珠宝商
机械代理人
制造业
纪念碑
音乐人
酒类零售店(酒精饮料)
干漆工
家禽养殖
房地产咨询
录音和播音员
饭店
屋顶建造承包人
制鞋业
写招牌的人
酒馆
出租车公司
轮胎商店
货车运输

食品设备
食品生产
铸造厂
葬礼神父
货车和拖拉机司机
室内装潢工人
批发市场
窗帘制造

C. 半职业性的

演员、杂耍演员
鉴定人员、估价师
军士长
商业性的艺术家
牧师(未经专业培训)
商店管理人员
州长代表
心理卫生学家
按摩人员
钢琴老师
公关人员
广播、电视播音员
法庭书记员
铁路调度员
室内装饰工
法庭翻译人员
实验室助手
景观设计师
殡仪业者
海军下级军官
新闻记者
检察官
标题搜索者
模具设计师
旅行代理
铁路调度工人

D. 农场人员

农场主

4. 文职人员和销售人员,技术员和小商行主

A. 职员和销售人员

银行职员和出纳员
账单回收者
簿记员
银行机器操作员
银行查账员
文职员或速记员
列车长
工厂仓库管理员
工厂管理员
邮局职员
路政人员
销售人员
军队士官、低级别军官
海运职员
共用事业工厂管理者
收费站管理者

B. 技术员

牙科医生
绘图员
驾驶教练
工厂发布文件的人
试验员
电话公司或工厂的教员
监督人员(重量、公共卫生、铁路、工厂)
调查员
专用分组交换机操作员
校对员
安全监督员
维护检查人员
技术助理
电话公司管理人员
计时员
铁路操作员

实验室技术员
机车维修技术员
车辆调度员
窗户清洁人员(商店)

C. 小商店所有者(3 000～6 000 美元)

花店店主
食品杂货店店主
报摊摊主
裁缝店店主

D. 农场人员

耕地所有人

5. 熟练的体力雇佣工

汽车修理工
面包师
理发师
锻工
装订工
锅炉(装配、修理)工
轫手,火车上控制煞车的煞车手
混凝土浇灌工人
干酪制造工
厨师
排字工人
制模工,开模工
柴油机发动机修理和维护人员(专业)
铲土机操作工
电工
雕刻师
杀虫剂喷洒人员
城市消防员
铁路消防员
煤气管道水管装配工
建筑业、奶品店工头
园丁(专业)
吹玻璃工人
装玻璃工人
军械工人
测量员
发型设计师
热处理工人
园艺家
公共用品的保养工人
油布、油毯铺设工人
啤酒制造者
推土机操作工
屠宰人员
橱柜制造者
电缆接线工
木匠
铸工、翻砂工人
技工(专业)
机器安装工人
制模工(专业)
油漆匠
裱糊工人
铁路巡逻人员
模具、模型制造工人
钢琴制造工
调音师
水管工人
城市警察
邮递员
印刷工
广播、电视维修人员
家庭用品修理工
绳绞接工
焊工(专业)
修船工
修鞋匠(专业)
文具设计(得到许可的)
服务员(俱乐部)
铁路转轨员
无轨列车操作员
工具制造工人

莱诺整行铸排机操作员
平版工
钳工
织布机维修工
机械工(专业)
维修工长
泥瓦匠
男按摩师
小农场人员
所有者
铁轨检察员
拖车引导员
印刷工人
装饰业者(专业)
钟表匠
织布工
焊接工
铁路调度场巡视人员
拥有农场固定财产的占有者

6. **机器操作者和半熟练雇佣者**

医院助工
学徒、电工、油漆工、锅炉工
装配工、精密工具制造者
装配线工人
酒吧间男招待
烈酒售卖者
桥梁看管、修护者
大厦管理员
公共汽车司机
检验员
自动贩卖机装填者
女管家
切肉工和包装工
读表工
工厂机器操作工
铁路加油工
有实际经验却无正式训练的护士
衣服熨烫工
抽水机操作工
收货人和检察人
盖屋顶的人
组织、策划人员(工厂)
冲锤、铣床工人
铁路信号员
工厂焊工
喷漆工
钢铁工(非专业)
快餐厨师
送货员
机器制衣工
电梯操作员
服兵役的士兵
文件编档员、磨砂机、缓冲器
铸造工人
修车厂和加油站服务员
花房工人
警卫
理发师
发电报的人
橡胶厂剥离工
出租车司机
鉴定员
计时员
轮胎制模工
铁路乘务员
普通卡车司机
男女服务生(较高级餐厅)
过秤员
现场焊接工
机器卷绕工人
拉丝工
酒类装瓶工
木工(机器)
商店和工厂的包装工

农场主
拥有极少资产的小房东

7. 无需技能的雇佣者

游乐园工人(保龄球场、赌注式台球)
除尘工人
停车场服务员
自助餐厅服务员
铁路洗车工
煤炭运输工
柜台服务员
牛奶厂工人
甲板人员
佣人
农场助手
渔民(拾蛤)
货物搬运工
捡垃圾者
挖墓者
砖头搬运工
屠宰人员
无特别要求的医院工人
铁路车辆维修工
看门人(清扫夫)
建筑工人
无特别要求的体力劳动者
洗衣店工人
邮递员
车站月台工人
小贩
守门人
公共或私人救济人员
盖或修屋顶的人
T恤衫折叠工
擦皮鞋者
把废品分类拣选的工人
舞台布置工人
码头装卸工人
货物管理员
马路清洁工
铁路压路工人
失业人员(没有职业)
无需技能的工厂工人
女服务生(普通餐馆)
洗车工
擦玻璃者
伐木工人

农场主
小佃户

7.2.6 教育量表

教育量表有一个前提假定,即接受过相似教育的男性和女性将会表现出相似的趣味和态度,同样也会表现出相似的行为模式。

教育量表被分成7个方面:

1)受过研究生专业培养:完成一项专业课程,并获得被社会认可的研究生学位的人,得分为1。

2)受过标准学院或全日制大学培养毕业:任何结束4年制学院或大学课程并得到受承认的大学学位的人被赋予同样的分数。州立大学和私立学院之间不做区分。

3)受过不完全大学培养:结束了至少一年大学学习但不是所有大学课程的人被

赋予这一位置。

4)高中毕业:所有高中毕业生被赋予这一分数,不管是毕业于私立预备学校、公立中学、中等专业学校还是教会学校。

5)不完全高中:结束了 10 年或者 11 年教育但是没有高中毕业的人被赋予这一位置。

6)初中毕业:结束了 7~9 年教育的人被赋予这一位置。

7)少于 7 年学校教育的:没有结束 7 年教育的人被赋予这一分数,不管他们接受的教育量有多少。

研究应用

Beck, Frances W. (1984, November). *Incidence of moderate and severe retardation.* Paper presented at the annual conference of the Mid-South Educational Research Association, New Orleans, LA.

Derogatis, Leonard R., et al. (1975). Social class, psychological disorder, and the nature of the psychopathologic indicator. *Journal of Consulting and Clinical Psychology*, 43, 183-191.

Evans, Charles S. (1982). Moral stage development and knowledge of Kohlberg's theory. *Journal of Experimental Education*, 51(1), 14-17.

Gray-Ray, Phyllis, & Ray, Melvin C. (1990). Juvenile delinquency in the black community. *Youth and Society*, 22, 67-84.

Healey, Gary W., & DeBlassie, Richard R. (1974). A comparison of Negro, Anglo, and Spanish American adolescents' self concepts. *Adolescence*, 33, 15-24.

Kelley, Eleanor, et al. (1974). Working-class adolescents' perceptions of the role of clothing in occupational life. *Adolescence*, 34, 185-198.

Morrison, J. W., Ispa, J., & Thornburg, K. (1994). African American college students' psychosocial development as related to care arrangements during infancy. *Journal of Black Psychology*, 20, 418-429.

Raph, Jane, et al. (1971, February). *Influences of a Piaget-oriented curriculum on intellectual functioning of lower-class kindergarten children.* Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

Rosenbaum, Sarah C. (1998). Gender differences in fear of power: The relationship between internal representations of mothers and attitudes toward authority. *Dissertation Abstracts International*, 59, 2431B.

Sherman, William Joseph. (1996). Role development in adult children of alcoholics: An experimental design. *Dissertation Abstracts International*, 57, 2164B.

7.3 评估群体的量表:结构与变化

这一部分包含了对五个量表的讨论,每一个量表测量了一个不同的变量,而这个变量是与群体的结构和变化有关的。

1)Hemphill 的群体维度指数,确定了群体的 13 个维度,是测量群体特征的最大胆

的尝试。

2)Bales 的互动过程分析是一个名义量表,广泛用于获取需要解决问题的群体中的个人互动的一些特征。

3)Seashore 的群体凝聚指数提供了对某一群体中群体力量的测量方法,而这种力量是用于维持该群体一致性和延续性的。

4)自发选择的社会人际量表和社会人际偏好揭示了群体成员之间的人际吸引。这些量表可以被修改,广泛应用于许多不同的情形。这不仅对于力图发现基本关系的研究者有用,同样对行为研究者或者社会工作者有用。围绕个人的新群体可以快速组成,而且关于群体的团队精神和劳动能力的测量方法也能够形成。

5)Bogardus 社会距离量表同样可以被修正以用于许多不同目的。两个人之间、一个人与一个群体之间或者群体之间的社会距离可以在许多不同的情形中测量得,比如那些群体外成员同一个国家、一个社区或者一个组织之间的距离。

7.3.1 Hemphill 的群体维度指数

被测变量 这一指数是被设计用来测量群体的特征维度的。

描述 这一指数建立在 13 个相对独立的群体维度之上:自主性、控制力、适应性、愉悦的氛围、同质性、亲密度、参与度、开放度、分化程度、群体实力、稳定性、分层以及凝聚力。根据 5 点量表回答 150 个项目。这些维度是从一些经常为权威所用的有关形容群体的词列表中选择的。所用的项目来自对 500 个人所做的开放式问卷的分析结果,然后由 5 个人将它们归入不同的维度种类中。

出处 John K. Hemphill 所著的《群体维度:测量手册》(*Group Dimensions: A Mannual for Their Measurement*,87 号研究专论;哥伦布市:俄亥俄州立大学,商业研究所,1956)。

信度 折半信度从 0.59 到 0.87。在预定项目上,某一选定的项目同最高类到最低类之间的相关度在 0.03 到 0.78 之间(中值为 0.36);而随机项目的则为 0.01 到 0.36(中值为 0.12)。维度指数之间的相关度在 -0.54 到 0.81 之间,大多在 +0.29 之内(显著性为 0.01)。不同报告者对于同一群体的一致性所测得的相关度在 0.53 到 0.74 之间。

效度 描述两个完全不同群体的特征的维度指数依群体不同而有变化,但是那些描述两个相似群体特征的维度指数则很相似。

效用 这一指数在用于研究领导者行为和它们所领导的群体特征之间的关系方面会有作用。尽管比较长,实施以及计算指数却相对容易一些。

问卷指导

根据你所描述的群体,将你对于各项目的答案写在答案纸上。不要在问卷上做任何标记。

在考虑每个项目的时候遵循下列步骤:

1)仔细阅读该项目。

2)考虑该题是否较好反映了你正在描述的群体的情况。

3)在答案纸上找到与你正在完成的题目的数字相对应的数字。

4)在答案纸上的每个数字后面你会发现五对虚线,以A、B、C、D、E标志。

- 如果你正在完成的题目所描述的该群体的情况是完全正确的,则将以A开头的那对虚线之间的空白处涂黑。
- 如果你正在完成的题目所描述的该群体的情况大多数是正确的,则将以B开头的那对虚线之间的空白处涂黑。
- 如果你正在完成的题目所描述的该群体的情况对错参半,或者你不确定是对还是错,则将以C开头的那对虚线之间的空白处涂黑。
- 如果你正在完成的题目在反映该群体的情况上大多是错的,则将以D开头的那对虚线之间的空白处涂黑。
- 如果你正在完成的题目在反映该群体方面是完全错误的,则将以E开头的那对虚线之间的空白处涂黑。

5)当涂黑两条线之间的空白处时,应该用浓黑线将所有空白处涂黑。如果你在涂答案的时候出错,则将整条线擦除并说明正确答案。

6)在极少数情况下,如果你认为一个项目根本不适用于该群体,或者你觉得你没有足够的知识来对这个项目所反映的群体的情况做出判断,则将这一题空着不回答。

7)结束一题后,依次往下做下一个。

你有足够长的时间来完成你的描述。请确认答案纸上的数字与问卷中所回答的题目的数字是对应的。

问　题

下面的问题使得我们能够客观描述社会群体的特定特征。这些项目只是描述群体的特征,他们并不判断这些特征是好的还是不良的。因此,这些问题绝对不应该被看做是对群体或者回答问题的人所做的测试。我们只是想要关于这一群体是什么样子的客观描述。

1. 该群体有为人熟知却又未成文,并与成员行为相关的规定。
2. 成员们不敢表达他们的真实观点。
3. 成员可能离开群体的惟一方式是被驱逐。
4. 如果一个成员想要离开群体,他不需要任何解释。
5. 如果一个人不能达到群体要求的话他将会失去成员资格。
6. 群体成员在严密监督下工作。
7. 群体内只有一些特定的观点可以自由表达。
8. 成员可以在任何他想的时间提出辞呈以离开该群体。
9. 成员离开群体的要求可能被拒绝。

10. 成员在群体会议上讲话的时候必须要三思。
11. 成员有时会被迫退出。
12. 群体成员遵守严格的纪律。
13. 群体规模迅速扩大。
14. 成员在不断离开该群体。
15. 群体内成员有较大的变动。
16. 成员们不断离开群体但是有新成员替代他们。
17. 在群体存在的整个期间内没有人离开。
18. 每个成员的个人生活被群体的其他成员所知。
19. 群体成员相互借钱。
20. 一名成员有机会认识群体内所有其他成员。
21. 成员们的关系没有密切到会产生相互之间喜欢或者不喜欢感觉的程度。
22. 群体成员相互之间帮一些小忙。
23. 所有成员之间很熟悉。
24. 群体内的每个成员都知道其他成员的名字。
25. 无论群体内外成员,都进行日常接触。
26. 群体成员是私人朋友。
27. 特定成员只在他们之间讨论私人问题。
28. 群体成员知道群体内其他成员的家庭背景。
29. 成员们相互之间直呼其名。
30. 群体由相互之间并不熟悉的个人组成。
31. 所有成员的观点都会被同等考虑。
32. 群体管理者比群体内其他成员地位高。
33. 群体内比较老的成员享有特权。
34. 群体受少数成员的行动支配。
35. 群体的每个成员享有同样的群体特权。
36. 有经验的成员控制群体。
37. 某些问题只在群体管理者之间进行讨论。
38. 某些成员比其他人对群体有更大影响力。
39. 群体的每个成员与其他任何人享有同样的权力。
40. 一个人在群体内的地位只取决于他做了多少事情。
41. 群体的某些成员在群体内有明确的职责。
42. 群体的元老级成员被给予特权。
43. 对群体的个人不满很小,不足以提出来。
44. 成会员们总是在抱怨他们为群体所做的事情。
45. 群体在完成工作的时候毫无活力和乐趣。
46. 失败的感觉笼罩着群体。
47. 在群体会议中经常有人大笑。

48. 该群体独立于其他群体而运作。
49. 群体有外部支持。
50. 该群体是一个更大群体的有效代表。
51. 群体活动受一个更大群体的影响,而它又属于这个更大的群体。
52. 群体外的人决定群体将要做什么工作。
53. 群体仿效其他群体所确立的榜样。
54. 该群体是组成一个大的组织的许多类似群体中的一个。
55. 群体所做的事情受到更高一级群体的允许。
56. 该群体与其他群体一起从事活动。
57. 该群体是一个更大群体的小部分。
58. 群体处于外部压力之下。
59. 成员受外部群体规范。
60. 群体的计划由位居其上的其他群体来制定。
61. 成员不允许从事任何影响群体发展的事情。
62. 成员以被接受为群体的一员而有一种荣耀感。
63. 该群体的成员资格是获取一般社会地位的方式。
64. 群体的失败对于个体成员不算什么。
65. 群体的活动占了每个成员醒着的时间的不足10%。
66. 成员们以加入该群体,从而在群体外人员中获取声望。
67. 群体的一名成员的错误可能导致所有成员陷入困境。
68. 群体的活动占据了每个成员醒着的时间的90%。
69. 该群体的成员资格可以帮助成员获取职业成就。
70. 群体的失败对大多数成员没有任何意义。
71. 如果群体失败,则每个成员都将失去自尊。
72. 该群体的成员资格给予每个成员一种高人一等的感觉。
73. 群体活动占据了每名成员醒着的时间的超过一半。
74. 群体的失败将会导致成员陷入困窘。
75. 成员不会因为为解决群体麻烦而作出的努力受到嘉奖。
76. 群体中有两或三名成员在遇到任何群体事情上总是站在同一边。
77. 某些成员对其他成员充满敌意。
78. 群体成员之间总是充满争执。
79. 成员知道每个人都在留意别人,同样留意自己。
80. 群体内的某些成员对其他人缺乏尊重。
81. 群体内的某些成员被认为是不合作的。
82. 群体内的某些部分总有一种相互暗中作对的趋势。
83. 群体成员作为一个整体共同工作。
84. 群体内的某些成员要为其他成员之间的争吵和一些敌意负责。
85. 亚群体之间存在能够影响群体活动的冲突。

86. 某些成员似乎没有能力作为群体的一部分而工作。
87. 成员之间有一种可能使群体分化的情感潜流。
88. 对群体有足够兴趣能参加群体会议的人被视作是成员。
89. 群体忙于增加成员。
90. 新成员在“越多越快乐”的基础上被欢迎加入群体。
91. 一个新成员只有在一个老成员退出后才可以加入。
92. 该群体的成员资格需要具备大学文化程度。
93. 一个人只要表达了加入的愿望就可以加入群体。
94. 任何想要进入群体的人都受到欢迎。
95. 成员资格对任何想要促进完成群体目标的人开放。
96. 准成员在进入群体之前要受到审查。
97. 任何申请群体资格的人都没有受到拒绝。
98. 要得到群体成员资格不需要特别的训练。
99. 是否能得到成员资格取决于个人受教育的程度。
100. 有兴趣加入群体的人被要求提供证书,这些证书将会受到检查。
101. 成员有较高的参与度。
102. 如果一个群体成员没有什么成就他将不会被鼓励留下来。
103. 群体的工作被留给那些被认为最有能力完成这项工作的人。
104. 成员对群体感兴趣但是有些人不想工作。
105. 群体有无所事事的名声。
106. 群体的任何成员都在一个或者更多有效的委员会中。
107. 群体的工作在成员之间有合理的分工。
108. 群体的每个成员都没有工作做。
109. 群体的运作经常因无事可做而中断。
110. 在一些比较长的时间内群体什么也不做。
111. 群体被引导向一个特定的目标。
112. 群体在多个目标中分派自己的精力。
113. 群体运作几套彼此冲突的计划。
114. 群体只有一个主要目标。
115. 群体明确知道它需要做什么。
116. 群体正同时致力于多个目标。
117. 群体做许多与它的主要目标没有直接关系的事情。
118. 群体的每个成员都清楚地知道群体的目标。
119. 群体的目标是明确的。
120. 某些成员忙于一件事情,另外的人则做别的事情。
121. 群体有一些在某种程度上相互冲突的主要目标。
122. 群体的目标从来没有被清楚地认识到。
123. 该群体是很不正式的。

124. 成员被要求遵守许多规章制度。
125. 群体在规定的有规律的时间内开会。
126. 该群体是沿着半军事化的层级组织起来的。
127. 群体的会议没有经过计划或者组织。
128. 该群体有一个组织表。
129. 该群体有指导群体行动的规范。
130. 该群体以组织表来安排人员。
131. 该群体有一张成员的姓名列表。
132. 群体的会议根据罗伯特秩序规则(robert's rules of order)来指导。
133. 关于群体活动有一个公认的展开行动的对或错的方式。
134. 群体面临的大多数事情需要投票表决。
135. 群体可以在任何碰巧合适的地方开会。
136. 群体成员的雄心壮志各不相同。
137. 群体成员来自同一个社会阶层。
138. 一些成员感兴趣于与其他成员完全不同的事情。
139. 群体包括一些背景存在广泛差异的成员。
140. 群体中有白人也有黑人。
141. 群体成员年龄都差不多。
142. 群体的少数成员比其他成员更有能力。
143. 群体成员代表了多种宗教信仰。
144. 群体成员在社会背景方面的差异很大。
145. 群体的所有成员都是同一种性别。
146. 群体成员的年龄差异至少有20岁。
147. 成员带着很不相同的家庭背景进入群体。
148. 群体成员的实际经验差别很大。
149. 成员在他们留在群体中的年数上有差异。
150. 群体包括不同种族的成员。

计分表和计分方法指导

关于某一特定维度的题目的得分是关于这一维度的题目的得分之和。比如,关于"控制"这一维度的原始得分为从第1到12项的分数之和。这一维度的总得分(原始分数)可能在12~60之间。

个别情况下,一个被调查者可能没有给出答案。对于这种遗失,计分为c(非对非错)。然而如果漏答的题目数量超过了分配给某一特定维度的项目的总数的一半,那这一维度就不要给分了。总地来说,经验表明几乎没有被访者故意漏答某个项目。

计分要素

控制力	A	B	C	D	E
1	5	4	3	2	1
2	5	4	3	2	1
3	5	4	3	2	1
4	1	2	3	4	5
5	5	4	3	2	1
6	5	4	3	2	1
7	5	4	3	2	1
8	1	2	3	4	5
9	5	4	3	2	1
10	5	4	3	2	1
11	5	4	3	2	1
12	5	4	3	2	1
稳定性	A	B	C	D	E
13	1	2	3	4	5
14	1	2	3	4	5
15	1	2	3	4	5
16	1	2	3	4	5
17	5	4	3	2	1
亲密度	A	B	C	D	E
18	5	4	3	2	1
19	5	4	3	2	1
20	5	4	3	2	1
21	1	2	3	4	5
22	5	4	3	2	1
23	5	4	3	2	1
24	5	4	3	2	1
25	5	4	3	2	1
26	5	4	3	2	1
27	5	4	3	2	1
28	5	4	3	2	1
29	5	4	3	2	1
30	1	2	3	4	5
分层	A	B	C	D	E
31	1	2	3	4	5
32	5	4	3	2	1
33	5	4	3	2	1
34	5	4	3	2	1
35	1	2	3	4	5
36	5	4	3	2	1
37	5	4	3	2	1
38	5	4	3	2	1
39	1	2	3	4	5
40	5	4	3	2	1
41	5	4	3	2	1
42	5	4	3	2	1
愉悦的氛围	A	B	C	D	E
43	5	4	3	2	1
44	1	2	3	4	5
45	1	2	3	4	5
46	1	2	3	4	5
47	5	4	3	2	1
自主性	A	B	C	D	E
48	5	4	3	2	1
49	1	2	3	4	5
50	1	2	3	4	5
51	1	2	3	4	5
52	1	2	3	4	5
53	1	2	3	4	5
54	1	2	3	4	5
55	1	2	3	4	5
56	1	2	3	4	5
57	1	2	3	4	5
58	1	2	3	4	5
59	1	2	3	4	5
60	1	2	3	4	5
群体实力	A	B	C	D	E
61	5	4	3	2	1
62	5	4	3	2	1
63	5	4	3	2	1
64	1	2	3	4	5
65	1	2	3	4	5
66	5	4	3	2	1
67	5	4	3	2	1
68	5	4	3	2	1
69	5	4	3	2	1
70	1	2	3	4	5
71	5	4	3	2	1
72	5	4	3	2	1
73	5	4	3	2	1
74	5	4	3	2	1
75	1	2	3	4	5

凝聚力	A	B	C	D	E
76	1	2	3	4	5
77	1	2	3	4	5
78	1	2	3	4	5
79	5	4	3	2	1
80	1	2	3	4	5
81	1	2	3	4	5
82	1	2	3	4	5
83	5	4	3	2	1
84	1	2	3	4	5
85	1	2	3	4	5
86	1	2	3	4	5
87	1	2	3	4	5
开放度	A	B	C	D	E
88	5	4	3	2	1
89	5	4	3	2	1
90	5	4	3	2	1
91	1	2	3	4	5
92	1	2	3	4	5
93	5	4	3	2	1
94	5	4	3	2	1
95	5	4	3	2	1
96	1	2	3	4	5
97	5	4	3	2	1
98	5	4	3	2	1
99	1	2	3	4	5
100	1	2	3	4	5
参与度	A	B	C	D	E
101	5	4	3	2	1
102	5	4	3	2	1
103	1	2	3	4	5
104	1	2	3	4	5
105	1	2	3	4	5
106	5	4	3	2	1
107	5	4	3	2	1
108	1	2	3	4	5
109	1	2	3	4	5
110	1	2	3	4	5
分化程度	A	B	C	D	E
111	5	4	3	2	1
112	1	2	3	4	5
113	1	2	3	4	5
114	5	4	3	2	1
115	5	4	3	2	1
116	1	2	3	4	5
117	1	2	3	4	5
118	5	4	3	2	1
119	5	4	3	2	1
120	1	2	3	4	5
121	1	2	3	4	5
122	1	2	3	4	5
适应性	A	B	C	D	E
123	5	4	3	2	1
124	1	2	3	4	5
125	1	2	3	4	5
126	1	2	3	4	5
127	5	4	3	2	1
128	1	2	3	4	5
129	1	2	3	4	5
130	1	2	3	4	5
131	1	2	3	4	5
132	1	2	3	4	5
133	1	2	3	4	5
134	1	2	3	4	5
135	5	4	3	2	1
同质性	A	B	C	D	E
136	5	4	3	2	1
137	1	2	3	4	5
138	1	2	3	4	5
139	1	2	3	4	5
140	1	2	3	4	5
141	5	4	3	2	1
142	1	2	3	4	5
143	1	2	3	4	5
144	1	2	3	4	5
145	5	4	3	2	1
146	1	2	3	4	5
147	1	2	3	4	5
148	1	2	3	4	5
149	1	2	3	4	5
150	1	2	3	4	5

群体维度概略与封面

姓名__________年龄________日期________

群体名称____________________

您的成员资格时间__________群体成员数量________

群体的一般性目的____________________

纬度		9级计分制分数								
		1	2	3	4	5	6	7	8	9
A	自主性	·	·	·	·	·	·	·	·	·
B	控制力	·	·	·	·	·	·	·	·	·
C	适应性	·	·	·	·	·	·	·	·	·
D	愉悦的氛围	·	·	·	·	·	·	·	·	·
E	同质感	·	·	·	·	·	·	·	·	·
F	亲密度	·	·	·	·	·	·	·	·	·
G	参与度	·	·	·	·	·	·	·	·	·
H	开放度	·	·	·	·	·	·	·	·	·
I	分化程度	·	·	·	·	·	·	·	·	·
J	群体实力	·	·	·	·	·	·	·	·	·
K	稳定性	·	·	·	·	·	·	·	·	·
L	分层	·	·	·	·	·	·	·	·	·
M	凝聚力	·	·	·	·	·	·	·	·	·

研究应用

Hemphill, John K. (1949). *Situational factors in leadership* (Bureau of Educational Research Monograph No. 32). Columbus: Ohio State University.

Hemphill, John K. (1956). *Group dimensions: A manual for their measurement* (Research Monograph No. 87). Columbus: Ohio State University, Bureau of Business Research.

Hemphill, John K., & Westie, Charles M. (1950). The measurement of group dimensions. *Journal of Psychology*, 29, 325-342.

Pheysey, Diana C., & Payne, Roy L. (1970). The Hemphill Group Dimensions Description Questionnaire: A British industrial application. *Human Relations*, 23, 473-497.

7.3.2 Bales 的互动过程分析

被测变量 群体互动。

描述 这一指数包含 12 个类别——表现出团结性、表现出缓解紧张、同意、提出建议、发表观点、给出指导、寻求指导、征求观点、征求建议、不同意、表现出不安、表现出敌意。通过给群体内的每个人分配一个分数而得到最后的分数。所有的互动都根据种类来分析并在互动发生时以 1—5 或者 1—0 的式样标注。通过观察,我们得到了

群体的摘要或总结，并且通过推断来描述群体基本的运行方式。

Bales(1970)发展了一个关于这些种类的略加修正的版本。围绕着"向上的/向下的，向前的/向后的，积极的/消极的"等维度，Bales建立了一个新的人际行为的等级体系。种类1现在指"看起来友好的"，而种类12表示"看起来不友好的"，种类2现在指"戏剧化夸张表现"，种类6和种类7分别表示"给予信息"和"寻求信息"，其他种类(除种类3、8和10外)的内容也都有所改变。

虽然现在Bales的方案并没有以其原本的形式被广泛使用，但它仍在其领域内保持了一种模型，这个方法对小群体分析的早期发展起到了很大的帮助作用。

出处 R.. F. Bales,《互动过程分析：一种小群体研究的方法》(*Interaction Process Analysis: A Method for the Study of Small Groups*. Cambridge, MA: Addison-Wesley, 1950)。

信度 当观察员素质较高并受过培训时，不同观察结果的相关度可以达到0.75～0.95。

效度 表面效度。参阅Lake, Miles和Earle的评论。

效用 这是一套通用的、标准化的种类集，非常适用于对小群体的观察和分析。其主要的缺点是对观察员的培训需要大量的练习。频繁的再训练也是必须的。

1	表示团结性：提高他人的地位，给予帮助、回报					
2	表示缓解紧张：开玩笑、笑、表示满意					
3	同意：表现被动的接受、理解、同意、顺从					
4	给予他人建议：指导、暗示他人自治					
5	发表观点：评价、分析、表达感觉、希望					
6	给出指导：信息、重复、阐明、确认					
7	寻求指导：信息、重复、证实					
8	寻求意见：评价、分析、感觉表达					
9	寻求建议：指导行动的可能方法					
10	不同意：显示被动的拒绝、遵循规范、拒给帮助					
11	表现出不安：寻求帮助、退出"场地(自己归属的群体)"					
12	表现出对立性：削弱他人的地位、自我防卫或坚持自己权利					

参考文献

Bales, R. F (1970). *Personality and interpersonal behavior.* New York: Holt, Rinehart & Winston.

Lake, Dale G., Miles, Mathew B., & Earle, Ralph B. (1973). *Measuring human behavior.* New York: Teachers College Press.

研究应用

Allen, William R., Comerford, Robert A., & Ruhe, John A. (1989). Factor analytic study of Bales' Interaction Process Analysis. *Educational and Psychological Measurement*, 49, 701-707.

Bell, Lorna. (2001). Patterns of interaction in multidisciplinary child protection teams in New Jersey. *Child Abuse and Neglect*, 25(1), 65-80.

De Grada, Eraldo, Kruglanski, Arie W., Mannetti, Lucia, & Pierro, Antonio. (1999). Motivated cognition and group interaction: Need for closure affects the contents and processes of collective negotiations. *Journal of Experimental Social Psychology*, 35, 346-365.

Jacobs, Janis E., & Bennett, M. A. (1993). Family decision-making during the transition into adolescence. *Journal of Early Adolescence*, 13,245-266.

Martinez, I. M., Orengo, V., & Prieto, Fernando. (1998). Development phases of task groups in videoconference communication. *Revista de Psicologia Social Aplicada*, 8(2), 91-107.

Mpofu, D. J. S., Lanphear, J., Stewart, T., Das, M., Ridding, P., & Dunn, E. (1998). Facility with the English language and problem-based learning group interaction: Findings from an Arabic setting. *Medical Education*, 32,479-485.

Pierro, Antonio, Sensales, Gilda, & Perez, Michela. (1998). Role differentiation in small groups: The effects of the sex composition of the group on expressive and instrumental behaviors. *Rassegna di Psicologia*, 15(1), 153-162.

Sells, James Nathan. (1995). The relationship between supervisor and traince gender and their interactional behavior. *Dissertation Abstracts International*, 56, 874A.

7.3.3 Seashore 的群体凝聚力指数

被测变量 该指数测量群体凝聚力,群体凝聚力的定义为群体的吸引力或对脱离群体的抵抗力。

描述 这个测试由三个问题组成:"你觉得你确实是你的工作团队的一部分吗?","如果你有机会去另一个工作团队做和现在相同的工作,获相同的报酬,你会过去吗?","在以下几点上,你的工作团队同中西部的其他工作团队相比怎么样?"最后一个问题包括三个方面的比较:成员之间共同合作的情况,成员相互团结的情况,成员在工作上互相帮助的情况。前面两个问题可以分为五个等级,第三个问题的三个方面可以分为四个等级。

出处 Stanley E. Seashore,《工业工作团队中的群体凝聚力》(*Group Cohesiveness in the Industrial Work Group*,密歇根大学社会研究所调查研究中心,1954)。

信度 构成凝聚力指数的量表所测量的各个群体的平均值之间的相关系数在 0.15 ~0.70 之间。

效度 该量表所测量的各群体显著性水平在 0.001 以上,这表明存在明显差异。

效用 由于问题是用短语表达的,该指数特别适用于比较机械的环境。经过部分修改后,它或许也适用于几乎所有需要群体凝聚力指数的情况。做这个测试只需花费极少的时间。应该向受访者保证,他们的回答会被予以保密。

你是否感觉到你真正是你们工作团队之一分子？

- □ 的确是
- □ 大多情况下是
- □ 有些情况下是，但有些情况下不是
- □ 我并未感觉到我真正属于这个团队
- □ 我不与团队的任何人一起工作
- □ 不确定

如果在其他团队中，做同样的工作，并获取同样的薪水，你是否考虑离开？

- □ 非常愿意离开
- □ 与呆在这儿相比，更愿意离开
- □ 没有什么差别
- □ 与离开相比，更愿呆在这儿
- □ 非常愿意呆在这儿
- □ 不确定

在以下几点上，你的工作团队同中西部的其他工作团队相比怎么样？

	比大多数要好	和大多数一样	没有大多数好	不确定
成员之间和睦相处的情况	□	□	□	□
成员相互团结的情况	□	□	□	□
成员在工作上相互帮助的情况	□	□	□	□

7.3.4 自发选择和社会人际关系偏好的社会人际关系量表

被测变量 个体在群体中被接受的程度，成员间的相互关系，以及群体的结构。

描述 对于紧密团结的小群体来说，测量结果是最令人满意的。社交测量技术要求群体中的每一个成员指出在进行一些特定的活动或处在某些特定情形下时，他们更喜欢同群体中的哪些人在一起。社交测量标准（选定的领域应当包含群体成员间可能发生的各种交往方式：工作、娱乐、拜访）从 1 个到 8 个或更多之间变化，回答时选择的人数可从 1 个到多个不等，由研究者的设计而定。

出处 J. L. Moreno，《谁将幸免？一种探讨人类关系问题的新方式》（*Who Shall Survive? A New Approach to the Problem of Human Relationships.* Beacon，NY：Beacon House，1934）。

信度 Loeb 的社交测量法的相关系数变化范围为 0.53 ~ 0.85，再测信度（test-retest reliablity）$r = 0.74$。

效度 Eugene Byrd（1946）对社交测量法中的选择情况和实际的选择情况的比较，以及间隔 8 周后的复测显示 $r = 0.76, 0.80, 0.89$。Gronlund（1951）在关于教师们

的实际看法和测试结果的对比显示:$r=0.59$。

参考文献

Byrd, Eugene. (1946). A study of validity and constancy of choices in a sociometric test. *Sociometry*, 9(2-3), 21.

Gronlund, N. (1951). *Accuracies of teachers' judgments concerning the sociometric status of sixth grade pupils* (Sociometry Monograph No. 25). Beacon, NY: Beacon House.

研究应用

Andrews, David, & Krantz, Murray. (1982). The effects of reinforced cooperative experience on friendship patterns of preschool children. *Journal of Genetic Psychology*, 140(2), 197-205.

Bukowski, William M., Hoza, Betsy, & Newcomb, Andrew F. (1994). Using rating scale and nomination techniques to measure friendship and popularity. *Journal of Social and Personal Relationships*, 11, 485-488.

Fjeld, Stanton P. (1965). A longitudinal study of sociometric choice and the communication of values. *Journal of Social Psychology*, 66, 297-306.

Herz, Joan A., et al. (1983, April). *Preschool children's concepts of their peers.* Paper presented at the biennial meeting of the Society for Research in Child Development, Detroit, MI.

Maras, Pam, & Brown, Rupert. (2000). Effects of different forms of school contact on children's attitudes toward disabled and non-disabled peers. *British Journal of Educational Psychology*, 70, 337-351.

Pettit, Gregory S., Bakshi, Anuradha, Dodge, Kenneth A., & Coie, John D. (1990). The emergence of social dominance in young boys' play groups: Developmental differences and behavioral correlates. *Developmental Psychology*, 26, 1,017-1,025.

Steiner, Ivan D., & Field, William L. (1960). Role assignment and interpersonal influence. *Journal of Abnormal and Social Psychology*, 61, 239-245.

Tagiuri, Renato, & Kogan, Nathan. (1957). The visibility of interpersonal preferences. *Human Relations*, 10, 385-390.

Tripathi, R. B. (1970). A sociometric study for the selection of section commander in "SICP" cadets. *Indian Psychological Review*, 6(2), 104-106.

Wait, Robert F., & Burke, Peter J. (1984). *Task performance factors and social-emotional leadership in small groups.* Paper presented at the meeting of the North Central Sociological Association.

(1)自发选择测试

对应每个名字,检测你对群体内成员的看法。

	喜欢	不喜欢	无所谓
Mary I.			
James F.			
John I.			
Sam E.			
Etc.			

(2)社交偏好测试

选出你最喜欢同其一起工作的5个人,[1] 标注出第一、第二、第三、第四、第五。

Mary I.			
James F.			
John I.			
Sam E.			
Etc.			

注:1. 其中有很多标准可用。比如,同在一个讨论小组中,作为你的邻居,一起玩桥牌,共同致力于某个项目,等等。

7.3.5 Bogardus 社会距离量表

被测变量 存在于特定的个人和某些社会群体之间的社会距离或社会接受程度。该量表适用于测量两个个体之间或两个或两个以上的社会群体之间的社会距离。这种方法已经被应用于种族距离、区域距离、性别距离、年龄距离、亲子距离、教育距离、阶层距离、职业距离、宗教距离、国家间距离等多方面的研究。

描述 典型的操作方法是,一群人被要求根据7个陈述将与社会距离程度相关的一组社会类型进行排序。在以"通过婚姻与之成为近亲"开始,直至以"应被驱逐出我国"结束。一个由100个担任评判的人组成的群体,从60个陈述中辨选出这7个陈述,按顺序排列成社会距离的连续统一体。

出处 Emory S. Bogardus,《社会距离》(*Social Distance*. Yellow Springs,OH: Antioch,1959)。

信度 折半信度系数达到0.90或更高。

效度 Theodore Newcomb 指出如果与在某些方面更为确切的其他量表配合起来使用,会获得高的效度。我们提倡在检验效度时运用已知群体(known group)方法。它包括找到那些以倾向于支持某些种族而不接受其他种族而闻名的群体。如果这些群体对量表的回答结果同基本的模式相符,则可以证明它的有效性。

计分 在这过程中,多种计分方法都曾被使用过。现在已发现有一种简单的方法与其他更复杂的方法同样可靠,即计算被检测到的最近的陈述的数目,也就是说,如果想计算许多人的种族距离商(racial distance quotient,RDQ),要先获得所有调查对象关于每个种族所选择的最近的陈述的总数目的算术平均数。如果要知道某一个人的RDQ,要先获得关于每个种族的最近的陈述的总数目的算术平均数。

效用 Bogardus 量表可以用于估算任何文化群体之间,在工业的、政治的、种族的、宗教的和社会生活等任何其他方面可能存在的同现实冲突的总量。它也有助于判断群体之间趋向冲突或趋向合作的程度。这种测量方法操作方便、计分容易,它也易于修改以适用于社会距离的其他问题的研究。

下面提供 Bogardus 种族距离量表的详细示例:

种族在这里主要被定义为一种文化的群体。

指导语:

1)请记住每一题都根据你的第一感觉反应填答。

2)请你把每个国家的所有人都看作一个整体。不要仅依据你对该国家人中最好的或最坏的人的印象做判断,而要以你对整个国家所有人的总体印象进行填答。

3)根据你的感觉指出你所能接受的类型,在每个国家人的后面划叉。

类　型	英国人	瑞典人	波兰人	韩国人	等等
1. 通过婚姻成为近亲					
2. 成为我在俱乐部中的私人好友					
3. 住在同一条街道成为邻居					
4. 成为我的同事					
5. 成为我国的公民					

研究应用

Brooks, Thomas Daniel. (2000). Differences in ethnic identity and social distance in White and Native Americans. *Dissertation Abstracts International*, 60, 3959A.

Fox, Louise W. (1984, April). *College students' attitudes towards social, physical and political deviancy*. Paper presented at the annual meeting of the Eastern Psychological Association, Baltimore, MD.

Howe, Esther. (1999). Reducing interethnic social distance among adolescents through participation in life-skills groups. *International Journal of Adolescent Medicine and Health*, 11, 411-428.

Kleg, Milton, & Yamamoto, Kaoru. (1995). Ethnic and racial social distance: Seven decades apart. *Psychological Reports*, 76(1), 65-66.

Kunz, Phillip R., & Yaw, Oheneba-sakyi. (1989). Social distance: A study of changing views of young Mormons toward Black individuals. *Psychological Reports*, 65(1), 195-200.

Lee, Sai Quon. (2000). Ethnic conflict: Social distance and trait attribution perspectives within four Southeast Asian ethnic groups in the United States. *Dissertation Abstracts International*, 60, 5836B.

Meyer-Lee, Elaine E. (1999). Understanding change in White adult students' racial identity and attitudes: The contribution of a developmental-contextual approach. *Dissertation Abstracts International*, 60, 3002B.

Oddou, Gary, & Clavijo, Fabio. (1983, August). *Teaching culture: The effects on attitude change and perceived similarity*. Paper presented at the annual convention of the American Psychological Association, Anaheim, CA.

Owen, Carolyn A., Eisner, Howard L., & McFaul, Thomas R. (1981). A half-century of social distance research: National replication of the Bogardus studies. *Sociology and Social Research*, 66 (1), 80-98.

Rapp, Adrian. (1983). Prejudice and discrimination among community college students. *Community College Social Science Journal*, 4(3), 63-66.

7.4 社会指标

社会指标一般被认为起着社会报告、社会系统说明和社会情报的角色,它是这样被定义的:社会指标——统计学、统计数列,以及所有其他形式的证据——是使政策制订者和其他决策制订者能够评定现行社会各个方面的运行情况,并对具体的规划进行评估,确定其影响的概括性的测量方法。社会指标有助于专家和外行人在社会变化的价值、目标和性质上,更好地理解他们自己的和其他的社会。斯图尔特·赖斯(Stuart Rice,1967 年)的研究对此有过简洁的陈述:

> 社会指标……为在纷繁的社会的各种相互联系中寻找出路所必需。它们描绘社会状态,彰显社会问题,预测社会趋势,通过社会工程,它们有望被导向社会计划所制定的社会目标。(p. 173)

社会指标的潜在范围非常宽广。在确定合适的范围之前应先解决目标问题,最好是根据社会的需要和政策制订者对他们必须碰到和解决的问题的信息的需求,来最终确定社会指标的范围。

参考文献

Rice, Stuart A. (1967). Social accounting and statistics for the Great Society. Public Administration Review, 27, 169-174.

7.4.1 社会指标的定义

无论社会指标是被应用于社区、州、国家,还是内政层面上,每个指标所指代的内容总是高度一致的。以下是一些用作社会指标的最常见的术语的汇编,改编自网页 www. socwatch. org. uy/indicators/Glossary. htm。改编行为征得了乌拉圭第三世界学会(the Third World Institute)的许可。释义部分中粗体字印刷的术语也在表单之中。

艾滋病 每 10 000 人中的艾滋病病例数。

面积 土地面积(平方公里)是指一个国家除去内陆水域部分面积之外的全部面积。大多数对内陆水域的定义都包括主要的河流和湖泊在内。

援助额占 GNP 百分比(捐赠国) 捐赠国支付的援助额占捐赠国 GNP 的百分比。

援助额占 GNP 百分比(受助国) 援助额占受助国 GNP 的百分比。

援助额(捐赠国) 捐赠国支付的援助额(百万美元)。

援助额(受助国) 受助国接受的援助金额(百万美元)。

人均援助额(受助国) 受助国人均接受的援助额(百万美元)。

对最不发达国家(LDC)的援助额(捐赠国) 发展援助委员会(DAC)向最不发达国家拨付的全部官方发展援助净值 (ODA),(百万美元)。

对最不发达国家的援助额(捐赠国) 发展援助委员会成员国向最不发达国家拨付的全部官方发展援助净值占捐赠国 GNP 百分比。

受到护理的分娩 受到医生、护士、助产士、受过正规训练的基础健康护理工作者或受过训练的传统接生人员护理下的分娩所占的百分比。

卡路里(Calories) 人均每日卡路里供应量表示一国每日食品净供给的卡路里当量与该国人口总量之比。

儿童死亡率(5 岁以下) 5 岁以下儿童死亡率表示每 1 000 个活产儿中 5 岁以下儿童的年死亡数量。更确切地说,它表示每 1 000 个活产儿在出生之后到 5 周岁之间死亡的概率。

避孕 避孕普及率是指自己或其性伙伴使用了任何形式的避孕措施的妇女所占的百分比。该指标一般只用于测量 15 ~49 岁之间的已婚妇女。

公民和政治权利公约 《公民和政治权利国际公约》,1966 年。

经济、社会和文化权利公约 《经济、社会和文化权利国际公约》,1966 年。

妇女政治权利公约 《妇女政治权利公约》,1953 年。

种族歧视公约 《消除一切形式种族歧视国际公约》,1969 年。

难民地位公约 《关于难民地位公约》,1954 年。

儿童权利公约 《儿童权利公约》,1989 年。

妇女歧视公约 《消除对妇女一切形式歧视公约》,1979 年。

国防开支(占 GNP 百分比) 军费开支占 GNP 百分比。北大西洋公约组织(NATO)国家的军费开支是基于北大西洋公约组织所制定的定义的基础之上,它包括国防部和其他部门与军事相关的支出(包括征募和训练军事人员,建造和购买军需品和军事装备)。国防部对于平民的花费不包括在内。军事援助包括在捐赠国的开支内,以赊购方式购买军事装备算在赊购当时而不是付款时。其他国家的数据通常包括国防部的开支(不包括公共秩序和安全方面开支,这些做另外归类)。

国防开支(占健康和教育开支的百分比) 军费开支占健康和教育开支的百分比。更多信息请看“国防开支(占 GNP 百分比)”。

国防开支(占全部开支的百分比) 军费开支占全部健康与教育开支的百分比。更多信息请看“国防开支(占 GNP 百分比)”。

森林砍伐 年森林砍伐率(1 000 公顷每年)。为转换成耕地、永久性的农业用地或居住用地而造成的永久性毁林。不包括诸如选择性采伐等其他方式的变动。

每 100 000 人的医生数 这里医生被定义为毕业于任何医学机构或医学院,现在正在国内任何医疗领域从事工作(实践、教学、研究)的人。

劳动收入份额(女性) 女性劳动收入占全部劳动收入的百分比。PNUD 计算基于以下估算数据:实际人均 GDP(PPP $),世界银行;从事经济活动人口比重,国际劳工组织(ILO);女性工资数占男性工资数百分比,国际劳工组织,联合国,以及 Psacharopoulos 和 Tzannatos。

劳动收入份额(男性) 男性劳动收入占全部劳动收入的百分比。更多信息请看“劳动收入份额(女性)”。

教育(开支) 用于教育事业的公共开支占GNP百分比。

教育开支(初等和中等) 用于初等和中等教育的公共开支占所有层次教育开支的百分比。

期望教育年限 儿童所被期望获得的平均正式受教育年数,包括大学教育和花费在复读上的年数。它也被认为是反映全部教育资源的指标,是用教育年限来衡量一个儿童在他或她的一生中需要呆在学校里的时间。男孩和女孩分开进行测量。

出口 出口(百万美元)。商品和服务的出口表示一国向世界其他地方供应的所有商品和其他市场服务的价值。包括商品、货运、保险、旅游和其他非要素服务的价值。劳动和财产收益(以前称为要素服务),例如投资收益、利息和劳动力收入等,不包括在内。所获得的数据以现值美元为单位。

外债(占GNP百分比) 外债总额是指一国以外汇、商品或服务对非本国居民偿还的债务。全部外债是指公众债务,公共担保的债务和未经担保的私人长期债务,国际货币基金组织的信贷支出,以及短期债务的总和。短期债务包括原始期限为1年或1年以内的全部债务和长期债务的余欠部分的利息。所获得的数据以现值美元为单位。

外债(百万美元) 详细信息请见“外债(占GNP百分比)”。

生育率 总生育率表示,根据特定年龄的生育率在每个年龄段所生育孩子的数量所得出的一个妇女在其整个育龄期内所生育孩子的数量。这一数据是观测值与估计值的结合,通常表达为“每个妇女的孩子数”。

外国直接投资 外国直接投资是指为获取在一个企业经营中的持续的管理权益(10%或以上的控股权)而进行的投资净流入量。它是按照一个经济体计算的,而不是按投资者计算的。它是国际收支中所列示的股份资本、收益再投资以及其他长期资本和短期资本之和。所获得的数据以现值美元为单位。

性别发展指数(GDI) GDI使用的变量和人类发展指数(HDI)使用的变量相同,但GDI根据男性和女性之间成就的不同调整了各个国家在预期寿命、受教育程度和收入水平等方面的平均成就。

性别赋权测量(GEM) GEM指数反映妇女能否积极参与经济和政治生活。该指数测量经济、政治和决策制订等关键领域的性别不平等情况,来集中关注妇女的参与机会。它因此也区别于反映基本能力的性别不平等情况的性别发展指数(GDI)。

性别赋权测量(排名) 根据性别赋权测量的指数值表明一国在世界排名中的位置。

基尼指数(Gini Index) 基尼指数是一种衡量收入或消费支出的实际分配情况偏离于假定的每人或每户完全平均分配的程度的综合性测量方法。基尼指数的最高值为100%,表示所有收入全部属于一个人或一户,最低值为0,表示绝对平等。洛伦兹曲线(lorenz curve)从最贫穷的个体或家庭开始,绘制出全部收入的累计百分数与接受救助性收入的累计百分数的比例。基尼指数测量洛伦兹曲线和预设的绝对平等线之间的面积,表示为:这个面积与绝对平等线下方最大面积的百分比数值。

国民生产总值(GNP)增长率 GNP年增长率以1987年本国货币固定价为基础计算。国民生产总值由本国居民创造的全部国内和国外价值总量构成,不计货币贬值因素的影响。它包括国内生产总值加上本国居民从国外获得的净要素收入,即向国外提供的有形服务(劳动力和资本)收入,减去支付给为本国经济做出贡献的非本国居民的类似款项。

人均国民生产总值(GNP) 表示人均GNP量。更多信息请见"国民生产总值(GNP)增长率"。

人均国民生产总值(GNP)增长率 表示人均GNP量的年增长率。更多信息请见"国民生产总值(GNP)增长率"。

医疗卫生(开支) 表示与国内生产总值(GDP)有关的医疗卫生开支总量。该指标被定义为用于医疗卫生开支的GDP份额。它包括所有公共的和私人的医疗卫生支出。

医疗卫生(服务) 该指标测量可以在1小时以内通过平常的交通方式到达就近的医疗卫生服务机构的人口比例。值得注意的是,卫生服务的便利设施往往都集中在城市地区。在某些国家,农村地区的医疗卫生服务水平要低得多。

人类发展指数(HDI) 作为一个复合指数,HDI根据人类发展的三个基本维度——长寿、知识和良好的生活标准——来测量一国的平均成就。用来表示这三个维度的变量分别是预期寿命、教育水平和实际人均国内生产总值。

人类发展指数(HDI)(排名) 根据HDI值显示一国在世界排名中的位置。更多信息请见"人类发展指数(HDI)"。

不识字率(成人) 成人不识字率是指15岁及以上的成人中不能读懂或写出关于其日常生活的简短陈述的人口的比例。

免疫(BCG) 表示根据国家免疫政策接种过BCG免疫抗原的适宜人口的比例。免疫的定义包括三个方面:①在1周岁之前接种过白喉、百日咳、破伤风混合疫苗(DPT)、麻疹、脑灰质炎、结核病、卡介苗、乙型肝炎疫苗的儿童比例;②非洲受影响国家中接种过黄热病免疫抗原的儿童比例;③接种过破伤风疫苗的育龄妇女比例。

免疫(DPT) 表示根据国家免疫政策接种过白喉、百日咳、破伤风混合疫苗的适宜人口的比例。详细信息请见"免疫(BCG)"。

免疫(麻疹) 表示根据国家免疫政策接种过麻疹疫苗的适宜人口的比例。详细信息请见"免疫(BCG)"。

免疫(脑灰质炎) 表示根据国家免疫政策接种过脑灰质炎疫苗的适宜人口的比例。详细信息请见"免疫(BCG)"。

进口 商品和服务的进口表示由世界其他国家提供给本国的所有商品和其他市场服务的价值。包括货物、商品运输、保险、旅游和其他非要素服务的价值。生产要素收入和财产收入(以前称为要素服务),例如投资收益、利息和劳动力收入等,不包括在内。所获得的数据以现值美元为单位。

收入(最高的20%) 最高20%的收入或者消费的百分比份额是根据收入或消费支出将人口按递增序列排序后,最高的五分之一的人口中所占有的收入或消费的份

额。有关个人或家庭的收入或消费的数据取自全国范围内有代表性的家庭户调查。在家庭户调查的原始数据可用的情况下,这些数据被用来直接计算五分之一人口的收入(或消费)分配。在其他情况下,分配情况通过能得到的最好的分组数据来估计。用于低收入或者中等收入经济实体的分布指标已经被修正以用于家庭,从而提供了一种测量人均收入或消费的更一致的方法。在测量国家内部生活成本的地理差异方面没有进行修正,因为用于这类计算的数据一般是无法获得的。由于下列家庭户调查在方法和收集的数据类型方面存在差异,分布指标在进行国家间比较时严格来讲是不可行的。随着调查方法的改进和日益标准化这些问题正在逐渐消失,但是具有严格的可比性仍然不可能。在任何可能的情况下,一般使用消费而不是收入数据。根据人均收入或消费情况对家庭进行排序以形成百分比序列,但这种百分比序列是以人口而非家庭为单位。关于低收入和中等收入经济实体的数据由世界银行政策研究部的贫困与人类资源分部编撰,采用了取自政府统计机构和世界银行各成员国的基本家庭户调查数据。关于高收入经济实体的数据取自国家性来源,以1990年卢森堡收入研究数据库、欧盟统计局统计年鉴以及联合国《国家会计统计:收入分布统计纲要》(*National Accounts Statistic: Compendium of Income Distribution Statistic*, 1985)加以补充。

收入(最低的20%) 最低20%的收入或者消费的百分比份额是根据收入或消费支出将人口按递增序列排序后,最低的五分之一的人口中所占有的收入或消费的份额。更多信息请见"收入(最高的20%)"。

收入(最高的20%和最低的20%之比) 收入最高的20%的家庭对收入最低的20%家庭的收入份额比,它是一种根据家庭总收入、人均收入(或支出数据)排序后,以递增序列划分为按百分等分排列(分布)的家庭分组,来测量收入或支出分配情况(或支出份额)的度量方法。在计算收入份额时使用了人口份额五等份和总收入或消费支出的最高十分位值。各国的数据主要取自参考时间段内在不同年份进行的全国范围内有代表性的家庭户调查。关于高收入OECD经济体的数据基于欧盟统计局(Eurostat)、卢森堡收入研究(the Luxemburg Income Study)以及经济合作与发展组织(OECD)的信息。由于不同的收入研究在利用收入和消费支出来估计生活水平上存在差异,在解释数据的时候应该慎重。

婴儿死亡率 婴儿死亡率指的是特定年份中每1 000个活产儿中在1岁以下死亡的婴儿数量。更准确地说,它是从出生到正好1周岁之间死亡的概率乘以1 000。这一数据是观测值与估计值的结合。一些国家如前苏联的一些成员,采用一种关于安全出生的非典型的定义,这一定义与世界卫生组织的标准定义相比减少了婴儿死亡率记录数。

劳动力 说明了劳动力人口的比例。总体劳动力包括那些符合国际劳工组织关于经济活动人口的定义的人:所有在一个特定时期内为商品或服务的生产提供劳动的人,包括在业人员和无业人员。尽管在对待诸如武装部队以及季节性或兼职工人等群体上不同国家之间存在区别,但一般而言劳动力包括武装部队、无业人员和初次求职者在内,而不包括家庭主妇和其他从事无报酬的护理者,以及在非正规部门工

作的人员。

劳动力(女性) 女性劳动力占总体的比例表明了女性在劳动力中活跃的程度。更多信息请见“劳动力”。

预期寿命 假设一个新生婴儿按照出生时的死亡率水平度过一生,那么这个孩子将会存活的年数。

低出生体重 出生时体重不足 2 500 克的儿童的比例。

疟疾 每 100 000 人中的疟疾病例数。

营养不良(5 岁以下) 5 岁以下儿童中低于相应年龄参照人口的体重平均值两个标准差以上的孩子的比例。

孕产妇死亡率 孕产妇死亡率指的是每 100 000 例生育中因怀孕或生产而导致死亡的妇女人数。由于生产时死亡在一些国家中被更广泛地定义至包含了由于怀孕或产后期间综合症或流产并发症而造成的死亡,而且许多孕期妇女死于缺乏适当的卫生护理,母亲死亡率难于在不同国家间进行一致而可靠的测量。显然许多母亲死亡案例没有被登记,特别是在一些有偏远农村人口的国家。这可能导致了一些过低的估计,特别是在几个非洲国家。孕产妇死亡率数据是官方估计数据,是来自行政机关的记录、在调查基础上进行的非直接估计以及根据世界卫生组织(WHO)和联合国儿童基金会(UNICEF)创建的人口模型而做出的估计。

官方发展援助(ODA) 由官方机构承担的给发展援助委员会(DAC)受援助国名单(发展中国家)第一部分的国家和地区的净赠款或贷款,以促进经济发展和社会福利为主要目标——并采取优惠的金融条件(如果是贷款,则至少其中 25% 属于捐赠)。ODA 净支付总额建立在经济合作与发展组织关于 DAC 成员国家、多边组织和阿拉伯国家数据的基础上。

每 100 000 人的护理人员人数 所有结束了基本护理教育的课程并且由国家给予资格证明、注册或授权,准许提供促进健康、预防疾病、照顾病患和康复等方面负责的、足够的服务的人数。

人口 总人口是基于“人口”的实际定义,所有居住者均计算在内,不论其法律地位或国籍如何。不在避难国永久定居的难民一般被认为是他们原属国人口的一部分。

人口(城市人口百分比) 居住在最近一次人口普查中官方定义的“城市地区”的人口所占的比例。农村百分比也同样计算。

贫困(国际线) 世界银行以每人每天收入 1 美元作为贫困线来进行国际比较。这一贫困线是根据商品和服务消费而来的。对于拉美国家和加勒比海国家建议贫困线为每天 2 美元。对于东欧国家和前苏联共和国,所采用的贫困线是每天 4 美元。在进行工业化国家之间的对比时,采用的是美国的贫困线,每天 14.4 美元。

贫困(国内线) 已经确立国内贫困线的发展中国家一般采用“食物贫困”的方法。这些贫困线表明,经济资源不足以满足基本最低食物需要。在工业化国家,国内贫困线一般被用来衡量相对贫困。欧盟委员会建议这些国家的贫困线为调整后个人收入的中间值的一半。

贫困(农村地区) 表明生活水平低于国内贫困线的农村人口的比例。更多信息请见

"贫困(国内线)"。

贫困(城市) 表明生活水平低于国内贫困线的城市人口的比例。更多信息请见"贫困(国内线)"。

受护理的怀孕妇女 表明怀孕期间受到护理的妇女的比例。

小学(读到5年级的孩子) 最终上完小学5年级的孩子所占比例(如果是4年制小学则是读到4年级)。这一估计是根据改造的同期群方法,使用连续两年的登记入学学生和留级学生的数据计算出来的。

每教师的学生数 小学学生—教师比是用小学入学的学生数除以小学教师的数量(不管他们的教学任务分别是什么)。

收音机 每1 000人拥有的用于接收针对一般大众广播的无线电接收设备的估计数量。

安全饮用水 获取安全用水是指能够获取到足量安全的饮用水(包括处理过的地表水和未经处理但是未遭污染的水,比如泉水、清水井和受保护的井眼)的人口比重。"足够"数量的水是用来满足新陈代谢、卫生及家庭需求的需要用水,通常一个人一天大约20升安全用水。"安全用水"的定义一直在随时间而改变。

卫生设施 获得卫生设施的人口比例是指享有至少足够的可以有效防止人、动物和昆虫排泄物接触的排泄物处理设施的人口比例。

学校入学率(小学,总体) 总体小学入学率。小学入学数据是对在小学登记的所有年龄的孩子占全国所有小学年龄阶段孩子的比例的估计。尽管许多国家认为小学年龄是6~11岁,但也有国家采用不同的年龄组。对于普及小学教育的国家而言总体入学率可能超过100%,因为一些学生的年龄可能高于或者低于国家标准小学年龄。

学校入学率(小学,净值) 净小学入学率。指的是在小学登记入学的达到官方规定接受初等教育年龄的人口的比例。这一指标被用于监测初等教育的参与程度并识别未入学的达到入学年龄的人口。接近100%的净入学率表明足够的初等教育容量的有效性以及积极的初等教育参与情况。低的净入学率表明在初等教育的普遍参与方面的不充分,原因在于缺少足够的学校或存在其他阻止孩子入学的因素。这一指标如果以性别来交互分类,则可以更显著地看出性别之间的差异。该指标对许多发达国家的适用性有限,因为在这些国家初等教育是强制性的,这样入学率一般接近甚至超过100%(因为许多超过通常年龄范围的学生也登记入学)。如果入学率比较低,通常表明是数据错误。

学校入学率(中学,总体) 总体中学入学率反映了中等教育的总注册人数占达到国家规定应该接受中学教育的年龄的总人口的比例。

学校入学率(中学,净值) 净中学入学率是达到官方规定的接受中等教育年龄的人口(按照国家标准)中实际进入中学的比例。

学校入学率(高等学校,总体) 高等学校总入学率是以在所有中学以上学校和全日制大学登记的学生数除以20~24岁年龄群体的总人口数而计算得到的。职业学校、成人教育项目、2年制社区学院和远程教育中心(主要是函授学校)的学生均包

含在内。这些学生在不同类型的学校里的分布情况因国家而异。将20~24岁的年轻人群体作为分母的定义已经被联合国教科文组织(UNESCO)接受,因为这一定义代表了高等教育学生群体的平均年龄水平,尽管高于或者低于这一年龄组的人也可能在高等学校注册。

科学家与技术人员 显示每1 000人中的科学家和技术人员的数量。“科学家”指的是在任何科学领域接受过科学或技术训练(通常完成高等教育)并专门从事研究和开发工作的科学家和工程师,包括那些领导研究和开发活动的行政管理人员和其他高层人员。“技术人员”指的是那些在结束中等教育的第一阶段后接受过至少3年职业或技术培训的从事科学研究和开发工作的人员。

中学入学率 显示了在中学登记入学的所有学生占按照国家规定属于中学教育年龄的总人口的比例。

社会支出 国家社会服务支出包括健康、教育、住房、福利、社会安全和社区服务设施等方面的支出。这些种类也包括:对由于生病和暂时性伤残造成的收入减少的补偿;支付给老年人、永久性残疾的人和失业者的费用;家庭、产妇及儿童补贴;福利服务的成本,比如照料老年人、残疾人和儿童。许多支出与环境保护相关,比如消除污染、水供应、卫生设施和垃圾收集都无差别地包含在这一类中。

电话 电话主线是将用户的设备与公用交换电话网络相连接的电话线路。所获得的数据是以全国每1 000人为单位的。

电视 每1 000人拥有的电视机数量。

高校学生(女性) 每100 000名女性中女大学生的数量。

失业 一定年龄以上,没有受到有薪雇佣或自雇佣的,但是可以获得并已经采取具体步骤寻找有薪雇佣或自雇佣的所有人。

失业救济支出 失业救济支出占整个政府支出的比例。

失业发生率(12个月) 长期失业的发生率(12个月以上),以百分比表示。

女性(薪水) 劳动力中每100名男性对应的女性。这一比例度量了在劳动力中男性和女性分别占的份额。

用母乳喂养孩子的妇女 用母乳喂养孩子至少6个月的妇女的比例。

年轻人失业率 失业的年轻人口占整个年轻劳动力的比例。更多信息请见“失业”。

研究应用

Boardman, Anthony P., Hodgson, Richard E., Lewis, Martyn, & Allen, Keith. (1997). Social indicators and the prediction of psychiatric admission in different diagnostic groups. *British Journal of Psychiatry*, 171, 457-462.

Bomalaski, Susan Huyler. (1999). Effects of counseling services with homeless women: Social indicators and intrapersonal outcomes. *Dissertation Abstracts International*, 60, 1842B.

Brown, Brett, Kirby, Gretchen, & Botsko, Christopher. (1997). *Social indicators of child and family well-being: A profile of six state systems* (Institute for Research on Poverty Special Report No. 72). Madison: Institute for Research on Poverty, Publications Department, University of Wisconsin-Madison.

Delgado, Melvin, & Barton, Keva. (1998). Murals in Latino communities: Social indicators of community strengths. *Social Work*, 43,346-356.

Herman-Stahl, Mindy, Wiesen, Christopher A., Flewelling, Robert L., Weimer, BeLinda J., Bray, Robert M., & Rachal, J. Valley. (2001). Using social indicators to estimate county-level substance use intervention and treatment needs. *Substance Use and Misuse*, 36,501-521.

Mammo, Abate, & French, John F. (1998). Using social indicators to predict addiction. *Substance Use and Misuse*, 33, 2 499-2 513.

Mazumdar, Krishna. (2000). Inter-country inequality in social indicators of development. *Social Indicators Research*, 49, 335-345.

Moore-Greene, Gracie. (2000). Standardizing social indicators to enhance medical case management. *Social Work in Health Care*, 30(3), 39-53.

Ryan, John A., Abdelrahman, A. I., French, John F., & Rodriguez, Gloria. (1999). Social indicators of substance abuse prevention: A need-based assessment. *Social Indicators Research*, 46(1), 23-60.

Zanders, Harry L. G. (1998). Social indicators. In Pieter J. D. Drenth, Henk Thierry, & Charles de Wolff (Eds.), *Handbook of work and organizational psychology. Vol. 2. Work psychology* (2nd ed., pp. 283-298). Hove, England: Psychology Press/Erlbaum.

7.4.2 人类发展指标

联合国开发计划署根据 Mark Malloch Brown 和 Richard Jolly 撰写的《1999 年人类发展报告》(1999 *Human Development Report*)给出了大量的数据摘要。下面所列的即该书的目录。

全球化与护理
照料与市场回报
再分配护理的费用与责任——给家庭、国家和公司
全球化经济中护理服务面临的挑战
第四章 使全球化促进人类发展的各国对策
抓住全球性机遇
保护人们免受伤害
克服资源挤压
使增长有利于穷人——减少不平等,增强人类潜在能力
建立国家角色的有效联盟
制定战略以解决全球体系中出现的新问题
第五章 再造全球治理结构——为了人道与公平
将人类利益和权利置于全球治理结构的中心
在经济危机中保护人类安全
减少危害人类安全的其他因素
缩小全球差距
采取特别行动以加强贫穷国家在全球治理结构中的协商地位
从现在做起建构21世纪所需要的全球体制
所有这些行动都始于人
特稿 与联合国建立伙伴关系,Ted Turner 撰写

7.4.2.1 人类发展指数

人类发展指数是可以在 www. undp. org/hdro/indicators. html #industrial 找到的众多不同的指标之一。这是一个由许多不同指标构成的评价总体生活质量的复合分数。表7.1给出了具有最高 HDI 的国家的排名。

7.4.2.2 联合国社会指标

联合国统计理事会创制了下列统计表作为社会指标。最新的数据可以在这一网址得到:www. un. org/Depts/unsd/social/index/htm。

(1)人口
估计人口数量(千人)
男性
女性
性别比例(每100名女性对应的男性数量)
平均年人口变化率(%)
(2)青年与老年人口
低于15岁的人口占全部人口的比例

60岁以上男性人口比例

60岁以上女性人口比例

60岁以上人口的性别比(每100名女性对应的男性数量)

(3)人类居住区

人口分布(城市人口对农村人口百分比)

平均年人口变化率(%)

城市

农村

(4)供水与卫生设施

能够获得较好饮水资源的人口比例

总体

城市

农村

能够获得较好卫生设施的人口比例

总体

城市

农村

(5)住房

平均每个房间中的人数

总体

城市

农村

(6)健康

出生时预期寿命

男性

女性

婴儿死亡率

儿童死亡率

男孩

女孩

(7)分娩

总和生育率

估计孕产妇死亡率

已婚育龄妇女避孕普及率(%)

任何避孕方法

现代避孕方法

(8)教育

预期学龄(预期受正规学校教育年数)

表 7.1 各国在不同人类发展指标上的排名

排名	国家	出生时预期寿命，1995年	成人识字率(%)	1995年整合的初等、中等和高等教育总入学比例	1995年实际人均GDP(PPP $)[a]	1995年调整过的实际人均GDP(PPP $)[a]	预期寿命指数，1995年	教育指数	人类发展指数，1995年
1	加拿大	79.1	99	100	21 916	6 230.98	0.900 8	0.993 3	0.96
2	法国	78.7	99	89	21 176	6 229.37	0.894 8	0.956 7	0.946
3	挪威	77.6	99	92	22 427	6 231.96	0.875 8	0.966 7	0.943
4	美国	76.4	99	96	26 977	6 259.29	0.856 2	0.98	0.943
5	冰岛	79.2	99	83	21 064	6 229.11	0.902 8	0.936 7	0.942
6	芬兰	76.4	99	97	18 547	6 218.88	0.856 3	0.983 3	0.942
7	荷兰	77.5	99	91	19 876	6 225.7	0.874 7	0.963 3	0.941
8	日本	79.7	99	78	21 930	6 231	0.914 2	0.92	0.94
9	新西兰	76.6	99	94	17 267	6 197.05	0.860 7	0.973 3	0.939
10	瑞典	78.4	99	82	19 297	6 223.42	0.889 5	0.933 3	0.936
11	西班牙	77.7	97.1	90	14 789	6 187.12	0.878 3	0.947 3	0.935
12	比利时	76.9	99	86	21 548	6 230.21	0.865 2	0.946 7	0.933
13	奥地利	76.7	99	87	21 322	6 229.71	0.861 7	0.95	0.933
14	英国	76.8	99	86	19 302	6 223.44	0.864	0.946 7	0.932
15	澳大利亚	78.2	99	79	19 632	6 224.81	0.886	0.923 3	0.932
16	瑞士	78.2	99	76	24 881	6 254.05	0.887	0.913 3	0.93
17	爱尔兰	76.4	99	88	17 590	6 198.1	0.856	0.953 3	0.93
18	丹麦	75.3	99	89	21 983	6 231.11	0.838 8	0.956 7	0.928

续表

排名	国家	出生时预期寿命，1995年	成人识字率(%)	1995年整合的初等、中等和高等教育总入学比例	1995年实际人均GDP(PPP $)[a]	1995年调整过的实际人均GDP(PPP $)[a]	预期寿命指数，1995年	教育指数	人类发展指数，1995年
19	德国	76.4	99	81	20 370	6 227.27	0.857 2	0.93	0.925
20	希腊	77.9	96.7	82	11 636	6 140.28	0.881 3	0.918	0.924
21	意大利	78	98.1	73	20 174	6 226.68	0.882 8	0.897 3	0.922
22	以色列	77.5	95	75	16 699	6 195.11	0.875 5	0.883 3	0.913
23	塞浦路斯	77.2	94	79	13 379	6 178.34	0.870 3	0.889 7	0.913
24	巴巴多斯	76	97.4	77	11 306	6 135.82	0.850 2	0.905 4	0.909
25	中国香港	79	92.2	67	22 950	6 232.88	0.900 7	0.837 9	0.909
26	卢森堡	76.1	99	58	34 004	6 286.9	0.851	0.853 3	0.9
27	马耳他	76.5	91	76	13 316	6 177.83	0.858 8	0.86	0.899
28	新加坡	77.1	91.1	68	22 604	6 232.28	0.868 7	0.832 5	0.896
29	安提瓜和巴布达	75	95	76	9 131	6 102.08	0.833 3	0.886 7	0.895
30	韩国	71.7	98	83	11 594	6 139.72	0.779	0.929 8	0.894
31	智利	75.1	95.2	73	9 930	6 115.55	0.835 5	0.876 5	0.893
32	巴哈马群岛	73.2	98.2	72	15 738	6 191.43	0.802 8	0.895 8	0.893
33	葡萄牙	74.8	89.6	81	12 674	6 171.35	0.830 5	0.867 3	0.892
34	哥斯达黎加	76.6	94.8	69	5 969	5 968.72	0.860 3	0.861 3	0.889
35	文莱	75.1	88.2	74	31 165	6 282.54	0.834 7	0.836 3	0.889
36	阿根廷	72.6	96.2	79	8 498	6 090.16	0.793 8	0.904 5	0.888

37	斯洛文尼亚	73.2	96	74	10 594	6 125.71	0.803 5	0.886 7	0.887
38	乌拉圭	72.7	97.3	76	6 854	6 048.8	0.794 3	0.901 6	0.885
39	捷克共和国	72.4	99	70	9 775	6 113.04	0.789 7	0.893 3	0.884
40	特立尼达岛和多巴哥岛	73.1	97.9	65	9 437	6 107.42	0.801 7	0.870 2	0.88
41	多米尼加	73	94	77	6 424	6 031.66	0.8	0.883 3	0.879
42	斯洛伐克	70.9	99	72	7 320	6 062.94	0.765 7	0.9	0.875
43	巴林	72.2	85.2	84	16 751	6 195.29	0.786 8	0.847 4	0.872
44	斐济	72.1	91.6	78	6 159	6 015.99	0.784 5	0.870 9	0.869
45	巴拿马	73.4	90.8	72	6 258	6 022.74	0.806 5	0.843 9	0.868
46	委内瑞拉	72.34	91.1	67	8 090	6 081.66	0.788 5	0.829	0.86
47	匈牙利	68.9	99	67	6 793	6 046.67	0.731 7	0.883 3	0.857
48	阿联酋	74.4	79.2	69	18 008	6 209.19	0.823 5	0.759 2	0.855
49	墨西哥	72.1	89.6	67	6 769	6 045.81	0.785 3	0.821 3	0.855
50	圣吉特与尼维斯岛	69	90	78	10 150	6 119	0.733 3	0.86	0.854
	平均	73.5	95.7	78.7	16 241	6 193	0.808 7	0.900 2	0.896

资料来源:该表中的数据可以在网页 www.undp.org/hdro/98hdl.htm 上找到,也可见 Mark Malloch Brown 和 Richard Jolly 的《1999 年人类发展报告》(纽约:牛津大学出版社)。

注:PPP $ = 购买力平价美元,一种反映不同国家间商品价格差异的调整数据。

总体

男性

女性

(9)读写能力

估计成年人口(15 岁以上)不识字率(%)

男性

女性

不识字率

15～24 岁(计算总体比例,男性和女性分开计算)

25 岁以上(计算总体比例,男性和女性分开计算)

(10)收入与经济活动

人均国内生产总值($美元)

成人经济活动人口比例(%)

男性

女性

失业率(失业人口占总劳动人口的比例)

总体

男性

女性

7.4.3 国家社会指标

下面列出的是美国联邦政府一级的各个部门使用的部分社会指标。

7.4.3.1 人口与家庭

18～34 岁的妻子的预期终生生育率,选定年份,分种族考察。

家庭生活的满意度以及对于和成年子女一起生活的老年人的态度,婚姻幸福以及离婚的方便。

(1)人口增长与构成

- 人口增长,以特定的年份作为标准
- 平均年人口变化率,以特定的年份作为标准
- 人口,分性别与年龄考察
- 人口,分种族和年龄考察
- 人口变动的组成,以特定的年份作为标准
- 生育率,以母亲的年龄区分考察
- 特定组群的累积生育率
- 接受的迁入人口,以主要职业分组,以特定的年份作为标准
- 平均年迁移率,以都市状况区分考察
- 在国内,与大城市及其郊区(SMSAs 美国行政区划单位)不相邻地区的年平均迁移率,以人口密度区分

(2)家庭规模与组成

- 家庭,以类型区分,以特定的年份作为标准
- 平均家庭规模,以成员年龄区分,以特定的年份作为标准
- 家庭,以类型和种族区分
- 丈夫不在的家庭中妇女的婚姻状况,以种族区分
- 家庭,以18岁以下的自己孩子的数量和种族区分
- 家庭,以规模和种族区分

(3)生活安排

- 人们在家庭中的地位,以年龄和性别区分
- 无亲缘关系的个人的生活安排,以年龄和性别区分
- 家庭中18岁以下的儿童,以父母是否健在和孩子的种族区分

(4)婚姻状况和稳定性

- 14岁及以上人口的婚姻状况,以性别区分,以特定的年份作为标准
- 14岁及以上的已婚妇女,以特定的年龄组、特定的年份区分作为标准
- 女性的初婚、离婚和再婚
- 配偶健在的已婚者中每1 000人中的离婚者的比例,以种族、性别和年龄区分,以特定的年份作为标准
- 涉及孩子的离婚判决,以涉及的孩子的数量区分
- 离婚和死亡率造成的婚姻解体
- 第一和第二次婚姻的离婚率,以性别和选定的组群区分

(5)国家间的比较

- 人口增长,以世界的不同地区区分
- 人口加倍所需要的年数,以世界不同地区区分
- 人口增长,以特定的国家作为标准
- 人口加倍所需要的年数,以特定的国家作为标准
- 选定的国家在选定的年份内的总和生育率
- 年龄增长与依赖性比率指数,以世界的不同地区区分
- 选定国家的年龄增长与依赖性比率指数,以特定的年份作为标准

7.4.3.2 健康与营养

(1)公众的看法

- 对健康的自我评价,以年龄和收入区分
- 选定的与健康领域有关的观点

(2)健康资源、利用与成本

- 初级护理医师与在业的牙医,以所处区域区分
- 去护理医师及牙医处看病,以性别、年龄和病人的家庭收入区分
- 从短期停留的医院出院以及平均住院时间,以年龄、性别和病人的家庭收入区分
- 心理健康机构的住院病人及出院病人护理经历
- 最近进入州及县心理医院的病人,以初诊为准,以病人的性别区分

- 接受出院病人心理健康服务的人，以初诊为准，以种族及病人的性别区分
- 国家及个人的健康护理支出，以付账的来源区分，以特定的年份作为标准
- 选定的医疗护理开支的消费者价格指数，以特定的年份作为标准

(3)生活机会

- 出生时预期寿命，以种族和性别区分
- 在20岁和65岁时的预期寿命，以种族和性别区分
- 由于心脏疾病及恶性肿瘤导致的50岁及以上的人的死亡率，以特定的年份作为标准
- 儿童和十余岁少年的死亡率，以特定的因素和年份作为标准
- 婴儿死亡率，以种族区分

(4)健康状况

- 伤残的天数，分不同类型，以性别和病人的家庭收入区分，以特定的年份作为标准
- 特定的慢性病的流行度，按特定的人口特征区分
- 患高血压的成年人，以性别、种族和年龄区分
- 在上一年做过血压检查的成年人，以性别、种族和年龄区分

(5)预防与营养

- 注射过麻疹、风疹、白喉、百日咳、破伤风混合疫苗和脑灰质炎免疫抗原的1~4岁儿童，以种族区分
- 吸烟的人，以性别、种族区分，特定的年龄组，以特定的年份作为标准
- 平均每天消耗0.501盎司或者更多纯酒精的18岁及以上的人，以特定的特征作为标准
- 被“皮肤褶”测量法测定为肥胖的20~74岁的人，以性别、种族、年龄和贫困状况区分
- 认为自己体重过重的17岁及以上的人，以性别区分
- 为减肥而节食的12~74岁的人，以特定的特征作为标准
- 人均铁摄入量，以性别、种族、贫困水平和年龄区分
- 出生前护理开始的时间，以3个月为一单位，以种族和母亲的年龄区分
- 高胆固醇的成年人，以年龄和性别区分
- 人均主食消费量

(6)国家间的比较

- 出生时预期生育率，以性别、选定的国家区分
- 婴儿死亡率，以特定的国家作为标准

7.4.3.3 住房与环境

(1)公众的看法

- 对住宅单元的评价，分处所和使用期限
- 对邻居的评价，分处所和使用期限

(2)住房清单与设施

- 住房清单的变化，分地理区域和地区

- 所有住房和 1970 年以后建在大城市及其郊区中心城市(SMSA,美国行政区划单位)的住房的居住者的特定的特征,以使用年限区分
- 房屋占有的变化,分种族和地理区域
- 住房问题,选定的地区
- 房屋价格的上涨与家庭收入
- 信贷与非信贷住房的相对租住成本
- 在所拥有的住房内使用的家庭取暖燃料的类型,分地区,以特定的年份作为标准

(3)房屋需求的趋势

- 人口增长,以地区区分
- 住宅设计的需求,以住房单元类型和来源来区分

(4)社区服务与社区福利设施

- 不充分的社区服务,以使用期限和所处位置区分
- 动迁愿望,以不充分的社区服务、使用期限和所处位置区分
- 特定的令人不满的周边环境,以使用期限区分

(5)环境质量

- 特定的附近地区污染源,以使用期限区分
- 估计的空气污染排放量,以来源区分,以特定年份作为标准
- 估计的进入水道的污染物排放,以来源和地区区分
- 联邦政府在改善和保护环境方面的支出
- 对改善和保护环境的支出的评估

(6)国际比较

- 每栋房子的房间数量和每户的人数,特定的国家,最近的年份
- 每个房间的人数和每个房间平均超过 1.5 人的住房,特定的国家,最近的年份
- 具备特定的设施的住房,特定的国家,最近的年份

7.4.3.4 交 通

(1)公众的看法

- 不充分的特定的周边服务
- 令人不满的特定的与交通有关的周边条件
- 与去年相比对于上班所采用的交通方式的满意度
- 对于上班所采用的交通方式方面的变化的满意度
- 对于特定的交通方面关注点的看法
- 对特定的交通设施的状况及未来的需求

(2)交通在美国人生活中的作用

- 交通对于美国经济的重要性
- 相关经济与交通发展趋势
- 货运与交通支出,以特定的年份为标准

(3)市内交通

- 以交通工具类型区分的市内交通
- 能被公共交通工具所搭载的市内乘客
- 客车使用,以目的区分
- 汽车使用方面的变化
- 使用中的汽车,以使用年数区分,以特定的年份作为标准
- 机动车辆燃料消耗与每加仑燃料行驶英里数

(4)社会代价与影响

- 机动车英里里程与单位英里死亡率
- 个人交通消费支出
- 大气污染物,以类型区分
- 通常的大气污染物,以类型和来源区分

(5)交通能源

- 能源消耗,以最终使用部分区分,以特定的年份作为标准
- 用于交通的能量消耗
- 人均国内生产总值与人均能量消耗,以特定的年份作为标准
- 美国客车的燃料使用效率

(6)国家间比较

- 机动车登记数量,以特定的国家和年份作为标准
- 机动车事故率,以特定的国家作为标准

7.4.3.5 公共安全

(1)公众的看法

- 害怕独自夜行的人,以特定的年份作为标准
- 关于法庭对付犯罪行为的措施的看法,以特定的年份作为标准

(2)对财务的责任

- 警察保护活动的开支,以政府级别区分,以特定的年份作为标准
- 警察保护活动的人均支出,以州来区分
- 警察保护活动的人均支出,以特定的城市作为标准
- 全职警察的保卫雇员,以特定的城市作为标准
- 个人暴力及盗窃犯罪,以特定的城市作为标准

(3)犯罪及其受害者

- 报告给联邦调查局的暴力犯罪,以类型区分
- 报告给联邦调查局的财产犯罪,以类型区分
- 谋杀案的受害者,以种族和性别区分
- 个人暴力及盗窃犯罪,以类型区分
- 个人暴力及盗窃犯罪,以性别、种族和西班牙裔受害者为区分
- 个人暴力及盗窃犯罪,以受害者年龄区分

- 家庭犯罪,以受害者种族及犯罪类型区分
- 家庭犯罪,以受害者的种族、居住年限和犯罪类型区分
- 家庭犯罪,以受害者的家庭收入和犯罪类型区分
- 家庭犯罪,以种族、受害者的家庭收入和犯罪类型区分
- 报告给警方的受害事件,以犯罪类型区分
- 报告给警方的受害事件,以受害者的种族和犯罪类型区分

(4)其他公共安全指标

- 交通事故引起的死亡,以类型区分
- 火灾造成的死亡
- 火灾造成的死亡,以性别和种族区分
- 火灾造成的死亡和伤害,以年龄和性别区分

(5)国家间比较

- 因交通事故和自然因素造成的意外死亡,以特定的国家作为标准
- 因工业事故、谋杀和其他原因造成的意外死亡,以特定的国家作为标准
- 机动车事故造成的死亡,以特定的国家作为标准
- 谋杀造成的死亡,以特定的国家作为标准
- 火灾造成的死亡,以特定的国家作为标准

7.4.3.6 教育和培训

(1)公众的看法

- 父母对他们上小学的孩子教育的意向和期望
- 对于改进国家教育系统的费用支出层次的看法以及对运作这一系统的人的信心
- 大学教育的重要性,以种族、教育程度和受访者的收入水平区分

(2)对财务的责任

- 公立与私立学校,以层次区分,以特定的年份作为标准
- 公立与私立小学及中学雇佣的正式教师,以特定的年份作为标准
- 公立与私立小学和中学中的生师比
- 高等教育机构中常驻性指导职员
- 教育、健康及国防支出占国内生产总值的比例
- 政府及地方支出,以类型区分,以特定的年份作为标准
- 每个学生在小学和中学教育方面的支出,以特定的年份作为标准
- 公立或私立高等学校的教育支出,以税收来源区分,以特定的年份作为标准

(3)入学与教育程度

- 3~34 岁的人的入学情况,以学校层次区分
- 18 岁人口中的中学毕业率,以州区分
- 18~24 岁的人口中大学入学情况,以性别和种族区分
- 18~34 岁的人口中大学入学情况,以种族和民族区分

- 18～24 岁未独立于家庭的成员中大学入学情况，以家庭收入分类
- 25 岁及以上的人的教育程度，以性别和种族区分

(4)表现与成就

- 8～17 岁的未能按时升级的学生，以特定的年份作为标准
- 对 9 岁儿童的教育过程的全国性评价，以科目和特定的调查对象特征区分
- 9 岁、13 岁、17 岁学生的科学成就的变化，以训练的类型区分
- 25～34 岁的全年全职工人的收入中值，以性别和教育程度区分
- 学士学位获得者的不充分就业情况，以性别和选定的大学专业为区分

(5)成人教育与培训

- 17 岁及以上未上过全日制小学、中学或者大学的人参加成人教育的情况，以教育程度区分
- 17 岁及以上未上过全日制中学或大学的人参加成人教育的情况，以提供者的类型区分
- 基础及中等程度成人教育的参与者，以年龄、种族和民族区分

(6)国际对比

- 15～24 岁的人的全日制学校登记入学率，以特定的国家作为标准
- 教育支出占整个 GNP 的比重，以特定的国家作为标准

7.4.3.7 工　作

(1)公众的看法

- 工作的投入
- 特定的工作特征的相对重要性
- 总体工作满意度，以特定的被访者特征区分
- 工作满意度指标

(2)劳动力资源

- 工人，以性别区分；非工人，以年龄区分
- 每 100 名工人中对应的非工人，以年龄区分
- 人口与劳动力，以性别区分
- 劳动力总量，以性别和年龄区分
- 完成 8 年或更少学校教育的劳动力，以性别和种族区分
- 完成 4 年或更多大学教育的劳动力，以性别和种族区分
- 在职工人的教育程度，以性别、种族和职业区分
- 选定组群的劳动力参与情况，以性别和受教育年数区分
- 不在劳动力行列内的人，以对工作的渴望及未参加工作的原因区分

(3)就业与失业

- 目前有工作的人，以持续时间和性别区分，以特定的年份作为标准
- 在职工人的职业分布，以性别和种族区分
- 有劳动能力且丈夫健在的已婚妇女，以是否有孩子和孩子的年龄区分，以特定的年份作为标准

- 已婚男性的就业状况,以其他家庭成员的就业状况区分,以特定的年份作为标准
- 自雇佣者与民间机构文职人员
- 政府雇佣和由于政府的企业以及对于商品和服务的采购而产生的就业,选定的年份
- 16 岁及以上的人的失业情况,以持续时间为区分
- 16 ~34 岁的人的失业率,以选定的特征区分,以特定的年份作为标准
- 未入学的 16 ~19 岁的人的劳动能力及就业状况,以性别区分

(4)条件与工作质量

- 可支配的平均周收入,以产业区分
- 在职工人的新就业情况与职业流动率,以特定的特征区分
- 制造业中人工周转成本率(labor turnover rate)
- 服务 35 年及以上年限的员工,区分年龄、性别和种族
- 平均周工作时间
- 被损耗的劳动时间,区分原因
- 轮班制工作的员工
- 有两份或更多工作的人,区分性别、种族
- 非洲人种和西班牙裔员工,区分性别、职业
- 停工和赋闲时间

(5)国家间比较

- 特定国家的劳动力参与率,区分性别
- 特定国家的失业率
- 特定国家特定年份的就业,区分经济因素

7.4.3.8 社会保障和福利

(1)公众认知

- 对社会问题上的支出的态度

(2)对财务的责任

- 特定年份,为特定社会福利项目公共和私人的支出
- 特定年份,特定公共计划,为福利项目的人均支出
- 特定年份,除了工资以外的个人收入

(3)提供的保障/保护

- 特定年份,OASDHI(美国最重要的社会保险计划称作 OASDHI,OASDHI 代表老年、幸存者、伤残和健康保险)的现金资助,区分受益人种类
- 特定年份,65 岁以上年龄,获得 OASDHI 现金资助,或 SSI 付款资助,或同时获得这两项资助的人数。
- 特定年份,65 岁以上人群的收入来源,区分婚姻状况和性别
- 65 岁以上的 OASDHI 受益人收入的中位数,区分婚姻状况和性别
- 特定年份,医疗保险制支付给年老的人的医保
- 特定年份,享有社会保险的员工
- 特定年份,对因短期病痛而造成收入损失的保障
- 特定年份,受到指定员工受益计划扶助的工人

- 特定年份,私人养老金和延期利润分享方案的受益者和相关报导
- 特定年份,对私人养老金和延期利润分享方案的捐助
- 特定年份,特定公共扶助项目,为每一笔受领金平均每月的支出
- AFDC[(美国)对有子女家庭补助计划]为每个家庭平均每月的支出,区分州
- 特定年份,AFDC 的受益者,区分父亲的社会地位
- 特定年份,医疗支出(美国医疗补助制度)
- 名为 XX 的社会服务的受益者,区分种类

(4)国家间的比较

- 特定国家,为特定社会福利项目的支出
- 特定国家特定年份,社会保障中的养老金、工资和价格的变化

7.4.4 国家和地方社区水平的社会指标

社会指标的发展并没有忽视国家层面。社会福利救助、婴儿死亡率、犯罪、缺乏教育以及其他方面被作为有问题的领域引入。这些方面的统计表和国家收集的数据相似,如 7.4.3 所列,为需要社会行为介入的社会问题的解决提供了线索。

地方性社会指标和国家指标有相同的标准。他们都必须证明自身的可测性,寻找社会价值并且共享目标,以及有政策价值,并且能够适合某个解释模型(该模型能解释经验性联系着的变量和指标之间的关系)。

7.4.5 目录:关于普遍适用的和特殊的社会指标以及最近发展

上百家资料来源为社会指标提供信息,而且不同的组织也设计了不同的适合他们自己项目的指标体系,表 7.2 提供了 11 个这样的例子。

表 7.2 社会指标的来源

资料来源	网 址
《健康》,美国,2000 年	www. cdc. gov/nchs/products/pubs/pubd/hus/hus. htm
联合国统计处统计数据——社会指标	www. un. org/Depts/unsd/social/
世界银行的社会指标	www. ciesin. org/lw-kmn/guides/sid. html
世界银行数据表	www. ciesin. org/IC/wbank/wtables. html
社会指标网	www. ccsd. ca/soc_ind. html
每月经济和社会指标	www. aph. gov. au/library/pubs/mesi/
白宫网站的社会统计发布室	www. whitehouse. gov/fsbr/ssbr. html
国际社会学协会社会指标工作组	www. ucm. es/info/isa/wg06. htm
社会公正指标	www. aminso. com/index. html
社会指标测量中心	www. columbia. edu/cu/ssw/projects/surcent/
欧洲社会指标系统	www. gesis. org/en/social_monitoring/social_indicators/EU_Reporting/eusi. htm

7.4.6 社会指标网络新闻

国际生活质量问题研究协会出版了社会指标网络新闻 SINET(the Social Indicators Network News)(www. soc. duke. edu/dept/sinet/),该出版物规范的涵括了社会指标的价值和使用。

国家经济指标 7.5

经济形势是社会形势的分支,通过经济指标可以调查社会形势,国家政府和研究者会使用多样化的经济指标收集数据,下文给出的是一些典型的例子。

(1)总产出,收入和支出

- 国内生产总值
- 实际国内生产总值
- 国内生产总值暗含物价紧缩稳定因素
- 国内生产总值和相关价格测量:指数和百分比的变化
- 非金融性公司业务:产值、价格、成本与利润
- 国民收入
- 实际个人消费支出
- 个人收入来源
- 个人收入分配
- 农业收入
- 共同利益
- 国内实际私人投资额
- 实际私人各类固定投资
- 企业投资

(2)就业、失业和工资

- 劳动力状况
- 选择性失业率
- 失业的选择性测量方法和失业保险计划
- 非农业就业
- 私人非农产业平均每周工作时间,每小时工资,每周的工资
- 就业成本指标:私人企业
- 生产率和相关数据,商业部分

(3)生产和商业活动

- 工业生产和利用能力
- 工业生产:主要市场群体和选择性产品

- 新建设
- 新的私人住房和空房比率
- 商业销售额和存货:制造业生产和商业贸易
- 工厂货运量、存货和订单

(4)价格

- 生产者价格
- 消费者价格:所有城市消费者
- 生产者成品货物价格的变动
- 消费价格变动:所有城市消费者
- 农民接纳并能支付的价格

(5)货币、信贷和保险市场

- 货币储备和债务测量
- 货币储备构成
- 储备金总额和金融基础
- 所有商业银行的银行贷款
- 非农业和非金融业企业的基金来源与使用
- 消费者信用
- 利率和债权收益
- 普通股价格和收益

(6)联邦财政

- 联邦收入、支出和债务
- 按种类划分的联邦政府的收入和按功能划分的联邦政府的支出
- 联邦部门、国家收入的主要款项

(7)国际统计

- 工业产量和消费者价格:主要工业国家
- 美国在产品和服务业中的国际贸易
- 美国国际贸易

7.6 组织结构的测量

一些关于组织结构的测量应该从社会学研究的角度考虑:

1)很多结构特征和人际关系,既存在于一个组织内部,也在多个组织间交叉存在着。

2)近年来组织结构测量取得了长足的发展,但还有一些关键性不足。

3)组织研究中应用的标准化测量不足,标准化的缺乏阻碍了组织理论的发展,并且要求研究者在选择组织测量方法时需要具备高度的识别能力。

4)尽管结构测量和描述是有趣的过程并有自己的实践,但重要的研究问题集中在群体

组织的多种结构安排的相互关联之中。

5)组织的内部因素,如士气、决策制订等影响着组织的外部关系,这些外部关系比如同其他组织是协作、防卫还是竞争。结构关系中的相互关系会暴露组织作为一个集合体的特质。

6)关于最重要结构变量的观点逐渐趋向一致,这些变量包括组织规模、正式化和集权化,其他的则也许出于研究者的兴趣,但这几个变量是研究设计中不能忽视的。

7)研究者对其他许多变量都很感兴趣,包括旷工、行政人员、疏远感、自治、权力基础、通讯、复杂性、舆论、协作、分散、分配公平、效用、革新、机械化、动机、惯例化、满足、控制范围、专门化、继任。

8)在对这些变量的评定中存在两套不同的测量方案:一种代表制度性方法,它基于文本和资料;另一种基于调查,以问卷和访问提纲的使用为特征。

9)对规模、标准化和集权化这三个变量的测量,最重要的是分析这三个变量和其他被选择的变量之间的关系。比如社会心理测量(组织中的工作压力指数)将在下文7.6.4.2中讨论。

研究应用

Ashkenas, Ron, Jick, Todd, Ulrich, Dave, & Kerr, Steve. (1995). *The boundaryless organization: Breaking the chains of organizational structure.* San Francisco: Jossey-Bass.

Fritz, Robert. (1996). *Corporate tides: The inescapable laws of organizational structure.* San Francisco: Berrert-Koehler.

MacKenzie, Donald, & Rogers, Vernon. (1997). The full service school: A management and organizational structure for 21st century schools. *Community Education Journal*, 25(3-4), 9-11.

Oliver, Barry D. (1989). *SDA organizational structure: Past, present, and future.* Berrien Springs, MI: Andrews University Press.

Person, Ruth. (1994). Organizational structure at the crossroads. *Educational Record*, 75(3), 42-46.

Pollard, George. (1995). Job satisfaction among newsworkers: The influence of professionalism, perceptions of organizational structure, and social attributes. *Journalism and Mass Communication Quarterly*, 72, 682-697.

Sauer, Mavis Anne Cheney. (2001). The mindful conduit: Organizational structure, climate and individual characteristics related to stress, communication and decision processes. *Dissertation Abstracts International*, 61, 2807A.

Sciulli, Lisa M. (1998). *Innovations in the retail banking industry: The impact of organizational structure and environment on the adoption process.* New York: Garland.

Stanley, Kay, & Reiboldt, J. Max. (1999). *IPA management: Organizational structure and strategic planning.* New York: McGraw-Hill.

Underwood, James C., & Hammons, James O. (1999). Past, present, and future variations in community college organizational structure. *Community College Review*, 26(4), 39-60.

7.6.1 组织测量手册

在组织测量的设计和选择上最适用的指导是 David Wealleans 的著作《组织测量手

册》(*The Organizational Measurement Mannual*,Burlington, VT:Grower,2001),下文附上内容目录:

5. 鉴定过程测量
    - 产量测量
    - 辅助测量
    - 计量
    - 测量标准
    - 概要
6. 目标和客观性
    - 为什么要有目标
    - 目标体系的建设性目的
    - 单项和双重限制
    - 建立目标水平
    - 调适目标
    - 目标发展的进度
    - 活动的限制
    - 概要
7. 探测和记录
    - 数据收集
    - 资源
    - 记录频率
    - 数据收集和记录的任务
    - 报告的方法和类型
    - 图表
    - 信息和表现
    - 概要
8. 解释
    - 我们需要解释吗
    - 避免争议
    - 怎样解释测量
    - 对分析的帮助
    - 处理测量结果的准备
    - 概要
9. 处理
    - 何时处理
    - 怎么处理
    - 检查计划
    - 问题的解决
    - 保持有利部分
    - 调整测量和目标
    - 概要
10. 相关人群

7.6.2 组织规模

定义 组织规模是测量组织运行的量度,对组织规模的测量包括全体职员人数、资产数额、支出额。在组织研究中,组织规模常常被表达为员工的数量,其实员工数量并不是测量组织运行规模必要的最好的量度。一家公司可能很大,但因为高度的机械化它可

能只需要很少的员工,当然,员工数量作为一个仍在被使用的指标是使用最普遍的量度。

测量 对测量组织规模的研究者的建议是根据研究设计方案确定的。几个或是更多的组织会被涉及吗? 这些组织规模是小的,中等大小的,还是大的? 什么样的分类是有用的? 是来自科室的、部门的还是整个组织的? 因为要研究组织的本质,可能涉及哪些敏感资料,比如,健康、政府或工业方面的资料。哪些资源是可用的? 资金流动只有通过开支票,还是可以打电话解决? 为私人会面设立过专款吗?

组织并不是偶尔发布信息,它们按惯例或者按规律来公布信息,一些必需的数据也许可以从组织的年度报告中获取,研究者如果想获取的是关于整个企业的图表,应该首先查看年度报告。

工业组织和商业组织中的员工数量可以在以下卷宗中查到,营业额 100 万以上美国公司名录提供了全世界范围内 100 000 家主要的公立的和私人企业的重要的商业信息,包括现在的地址、财政和市场信息,以及董事经理职位部门的人员名单。

如果所需要的信息不方便公开,研究者可以给劳资关系部门经理或人事部门经理写信、致电,或登门拜访表达研究需要、研究动机、研究任务。这些管理人员可以提供在部门和科室分歧中哪些是需要的(研究者应能明确指出他们将如何处理那些对组织有用的数据)。另一个办法是加强同市场和公关部门联系,这两个部门对传播任何(客观的)关于公司企业历史和活动的信息都很有用。

在多数公司中,员工人事部门会存放个人档案,薪工管理部门也会保管工资册,越往后的记录可能越准确。当然,工资册(和人员)一年中经常变动,因此研究者多大程度上把这个估计进去是很重要的。

Hall 在他的著作里提供了研究规模和它的其他相关方面的概要(1992,pp. 112-139),概要指出,在关于组织形式的结论上,规模这个要素引发了争论和分歧。无论怎样,比较大的组织比小的组织更加专门化、标准化、正式化,在这一点上正逐渐达成一致结论。规模和其他结构尺度之间,如权力集中和工作流程的直线控制之间联系的缺乏是相当显著的。Hall、Haas 和 Johnson(1967 年)研究了北美 75 家组织的数据,报告了他们的结论和发现:

> 这些发现直接表明,组织规模既不暗含复杂性,也不暗含着正式化程度的意义。社会科学家在大型组织中的引导研究能够很好地置疑当下流行的假设:即被研究的组织应该是高度综合化和正式化的,社会科学家需要经验地检查每个组织当下的综合化和正式化程度。(p. 911)

在 Hall 的书中(1972),他对规模研究的观点可以推导出一些组织规模与科技、专业化、工作流程、管理元素、个体、组织和社会关系的结论(p. 138),最重要的结论表述如下:

- 组织规模很大程度上被技术或为组织使用的技术所改变。
- 管理元素与组织规模之间呈现出曲线联系:管理元素在组织规模扩大的时候却倾向于减少;但是在巨大型组织中,管理元素在组织规模扩大的时候倾向于扩大。
- 组织的大规模对其中的个体有冲击:更大压力、人格分裂的压力给许多员工带来

不适。这些消极的结果可以部分地由组织内的非正式友谊消解。

- 规模大也导致了一些组织控制、调和、信息通达方面的困难。但同时,也给组织带来掌控环境更大的力量,执行计划更多的资源,并且更少地依赖某些特殊个人。
- 大型组织中权力集中可能集中社会的权力,从而影响民主化进程。

参考文献

Hall, Richard H. (1972). *Organizations: Structures and process.* Englewood Cliffs, NJ: Prentice-Hall.

Hall, Richard H., J. Eugene Haas, & Johnson, Norman J. (1967). Organizational size, complexity, and formalization. *American Sociological Review*, 32,903-912.

7.6.3 正式化

定义 正式化代表组织中规范的应用。一些组织详细地描述每个工作的具体权利、责任、义务和应遵循的程序,从而保证职位上的每个人能与职位的需求一致。处罚系统用文字的形式清楚地写下来,公正地监控以防违反纪律。否则组织就会因为有漏洞的工作细则,而无法对工人的工作进行仔细的监控。

正式化的这两种不同尺度可以详细分解为:工作说明或者工作标准化程度,规范宽容度或者行为在标准之下准许的尺度和宽容度。

测量 Aiken 和 Hage 做过关于正式化的更深入研究(1966;Hage & Aiken,1967a,1970),此外还有 Richard Hall 和他的同事以及其他研究小组也做过此类研究。Aiken 和 Hage 主要基于传统的调查方式,Hall 和其他研究者更多的基于文献资料。两种方式都值得推荐,但是从节约的角度考虑 Aiken-Hage 测量方式更适合。

计算 下面 5 个量度由 14 个问题构成,工作规章制定(问题 1 ~ 5),规章遵守(问题 6 ~ 7),规章手册(问题 8),工作描述(问题 9),具体工作特征(问题 10 ~ 14)。这 14 个问题的答案选项被分成 4 个量度:1(绝对合适)到 4(绝对不适合)。每个回答者的 5 个正式化量表中每个量表都取平均数,平均数越大,正式化程度越高(平均数最大值是 4)。研究者从 5 个量表的平均数进行记录,每个回答者被根据社会地位分类,在第一个平均数的基础上,计算组织中,每类社会等级的每个量表的第二个平均数。社会地位由被调查人在组织中所处的阶层,以及部门还有专业工作的种类来定义。比如,如果一家中介公司专业职员由精神病学家和社会工作者组成,每一个都划分等级,这家公司可以划分 4 个社会等级:管理层的精神病医生、精神病医生、管理层的社会工作者、社会工作者。组织在(5 个量表中)每个量表中的得分是通过计算组织中所有社会等级在每个量表中的平均值的平均值而得到的。

出处 Michael Aiken & Jerald Hage(1966),Organizational Alienation,*American Sociological Review*,31,pp. 497-507. 量表和数据的可靠性由 James L. Price 检验,*Handbook of Organizational Measurement* (Lexington,MA:D. C. Health,1972),pp. 108-111。

Hage 和 Aiken 正式化研究详细目录

这些数据是依靠访谈收集到的,采用了14个问题。

说明:我列举一些有关你在××组织中的工作的陈述,可能符合你的情况也可能不符合,对我读的每个题项,请根据你和你所在组织的实际情况,用如下选项回答。

- 绝对适合
- 比较适合
- 有些适合
- 绝对不适合

	绝对适合	比较适合	有些适合	绝对不适合
1. 首先,我感到在许多事务上我可以自己做主				
2. 组织中个人能做出他自己的决定而不必与其他人协商				
3. 事情是如何进行的主要取决于做这件事的人				
4. 这里的人们大多数时候可以按他们喜欢的方式做事情				
5. 大部分人工作中有自己的规则				
6. 员工经常被检查是否违犯条例				
7. 员工觉得经常被观察是否遵循所有的条例				
8. 这里没有条例手册				
9. 我的工作有书面的完全手册				
10. 不论情况如何,我们都有既定的处理程序				
11. 每个人都有明确的工作				
12. 比较强调按固有的规定行事				
13. 组织保存每个人工作成绩的书面记录				
14. 我们一直遵守严格的操作规则				

资料来源:Michael Aiken & Jerald Hage (1966) ,"Organizational Alienation",*American Sociological Review*,31,497-507. James L. Price 缩编了这个研究的详细目录。

信度 本研究没有数据涉及可靠性问题。

效度 正式化与相互疏远相关。组织的正式化程度越高,工作中相互疏远的可能性越大。在这些工作被严格结构化的组织中,存在着对工作的不满。严格的执行规章和对工作的不满是紧密联系着的,当规章被严格执行,社会关系就被打乱了。可以在日常工作和规章手册,工作描述和工作特征之间发现重要的积极关联。

效用 大多数案例中,访谈可以控制在5分钟之内。

参考文献

Aiken, Michael, & Huge, Jerald. (1966). Organizational alienation. *American Sociological Review*, 31,497-507.

Hage, Jerald, & Aiken, Michael. (1967a). Program change and organizational properties. *American Journal of Sociology*, 72, 503-519.

Hage, Jerald, & Aiken, Michael. (1967b). Relationship of centralization in other structural properties. *Administrative Science Quarterly*, 12, 72-92.

Hage, Jerald, & Aiken, Michael. (1970). *Social change in complex organizations.* New York: Random House.

7.6.4 集中化

定义 集中化指的是组织中权力的集中程度。权力是每个组织内部的重要因素。权力的分配对于组织的成绩和组织成员的行为有重要意义。

对权力的一个重要的考虑就是它的分配方式。如果所有权力都为某个个人行使,那么这就是权力集中的最大化;如果所有权力为组织中所有成员平等的行使,那么这就是权力集中的最小化。大部分的组织处在两个极端之间。

在不同的权力分配方式中,行为者使用权力而影响到上级、同级同事和下属,关于权力分配方式和权力集中程度会发生出各种问题。这些问题一般发生在权力的分层、参与管理的程度、企业的民主程度、群体决策的制订、员工代表、集体契约、疏远感和组织冲突中。

测量 对大多数测量来说,集权化都是通过制度的方式来评定,通过相关资料和知情者,或者使用调查的方式,把问卷调查和访谈作为主要的工具。

研究应用

Bohannan, Paul. (1994). The rise and fall of centralization. *Human Organization*, 53,395-398.

Cullen, John B., & Perrewe, Pamela L. (1981). Decision making configurations: An alternative to the centralization/decentralization conceptualization. *Journal of Management*, 7(2), 89-103.

Dallago, Bruno, & Mittone, Luigi. (Eds.). (1996). *Economic institutions, markets and competition: Centralization and decentralization in the transformation of economic systems. Northampton, MA: Edward Elgar.*

Edelson, Paul J. (1995). Historical and cultural perspectives on centralization/decentralization in continuing education. *Continuing Higher Education Review*, 59(3), 143-156.

Hall, Richard H., Jiang, Shanhe, Loscocco, Karyn A., & Allen, John K. (1993). Ownership patterns and centralization: A China and U.S. comparison. *Sociological Forum*, 8, 595-608.

Hetherington, Robert W., & Hewa, Soma. (1998). The consequences of centralization for performance in departments of a multihospital system. *International Journal of Contemporary Sociology*, 35(1), 28-56.

Huber, G. P., Miller, C. C., & Glick, W. H. (1990). Developing more encompassing theories about organizations: The centralization-effectiveness relationship as an example. *Organization Science*, 1(1), 11-40.

Piankoff, Janette. (1999). The effects of centralization, formalization and innovation norms on organizational commitment. *Dissertation Abstracts International*, 60, 1898B.

Prechel, Harland. (1994). Economic crisis and the centralization of control over the managerial process: Corporate restructuring and neo-Fordist decision-making. *American Sociological Review*, 59, 723-745.

Van Buren, Mark Everett. (1996). Technology fulcrums: Information technology, the centralization of decision-making, and performance in commercial banking. *Dissertation Abstracts International*, 56, 4170A.

7.6.4.1 Aiken 和 Hage 关于决策制订和权力层级中个人参与的量表

实际参与指数的操作化问题如下所示:

1. 你经常参与招募新职员的决策吗?

——从不 ——偶尔 ——有时 ——经常 ——总是

2. 你经常参与提拔任何员工的决策吗?
3. 你经常参与是否采用新政策方针的决策吗?
4. 你经常参与是否采用新流程的决策吗?

权力等级量表中的问题如下:

1. 直到主管人批准方案才能采取行动。

——绝对不适用 ——不适用 ——适用 ——绝对适用

2. 每个意图作出他(她)自己决定的人都会迅速遭到阻止。
3. 即使小事情也不得不(必须)依赖上级作出最终决定。
4. 在几乎做任何事情之前都要询问我的老板。
5. 我作出的任何决定都必须得到老板的批准。

计算 两种决策类型具有不同的计算法。为(计算)实际参与指标,五个回答分别标记上1(最低参与程度)至5(最高参与程度)分。回答"从不"的得1分,回答"总是"的得5分。计算每个回答者对五个问题回答的平均分数,然后对每个回答者按照社会地位进行分类,再按组织内不同的社会地位来计算第二个平均值。根据Aiken和Hage的观点,职员在社会组织中的层次位置,或职业活动所在的部门和所属的类型确定。例如,如果一家公司的职员是由精神病医生和社会工作者组成,他们都分成两个等级,这个公司就可以划分四种社会地位:管理层的精神病医生、精神病医生、管理层的社会工作者和社会工作者。这个组织的分数由计算该团体内所有社会地位等级平均数的平均水平决定。

权力等级量表的计算法与实际参与指数的计算法类似。回答被分别标记数字1(绝对不适用)到4(绝对适用),类似于实际参与指数的计算法,权力等级量表中组织

的得分是建立在不同社会地位等级平均数之上的，而社会地位等级的平均值是建立在每个回答者的平均分基础之上的。

出处 M. Aiken & J. Hage (1968), Organizational Independence and Intraorganizational Structure, *Americal Social Review*. 33, pp. 912-930。已得美国社会学协会和作者的使用授权。

信度 Aiken 和 Hage 未提供相关数据。

效度 仅由少数顶层人员制订决策的组织依赖制度和严密的监管，以此确保员工行为持续的一致性。这些组织也以拥有为数较少的专业职员为特征。拥有经过严格训练的员工对规章制度的需要就会减少。Pennings 认为高度自治的组织倾向于具有非参与性的内部决策结构，自治程度越高，主管人员的掌控范围就越大。

参考文献

Pennings, Johannes (1973). Measures of organizational structures: A methodological note, *American Journal of sociology*, 79, 686-704)

7.6.4.2 组织内与工作相关的紧张/焦虑指数

被测变量 这个指数旨在测量个体因工作而产生的紧张/焦虑(感受)程度。

描述 该指数共包含 15 项陈述性说明，它们描述了作者认为是(感受到)冲突或模棱两可的(焦虑)症状。(这里)要求回答者利用 5 分位的 Likert 量表对他们被每种症状困扰的频率进行估测。

出处 Robert L. Kahn & Donald M. Wolfe, *Organizational Stress; Studies in Role Conflict and Ambiguity*(New York: John Wiley, 1964), pp. 424-425。

信度 虽然对全国 725 个受雇佣成年人样本中进行了相关分析，但是再测信度值未能得出。另外，对 53 位管理人员进行了深入调查。总体看来，平均内部项目的相关性集中在 0.70 左右。该样本的自变量间内部相关的矩阵值与在全国样本中发现的非常相近。

效度 调查采用了开放式问卷来收集与工作相关的焦虑的数据、内容和强度的信息。这些都与紧张指数紧密相关。发现了某些紧张/焦虑和满意(感受)之间的间接相关。

效用 实施测试所需时间估计在 15 分钟以内。这一标准同样适用于雇员和管理人员。它既是诊断工具也是测量指数。该指数对紧张/焦虑感受的诊断能力可能是它最重要的作用。

计分 回答者从 6 个备选答案中选择一个做出回答：从不苦恼，很少苦恼，有时苦恼，经常苦恼，一直苦恼和(各选项)均不适用。前 5 个回答分别应得到 1～5 分。回答者的总分是他或她回答的全部问题除了那些回答者选“(各选项)均不适用”的项目的平均数。得分应在 0～5 分之间。

组织内与工作相关的紧张/焦虑指数

在工作中,我们都偶尔会对某些事情感到烦心苦恼,我将列举出一系列有时会让人苦恼的情况,并请告诉我你对每项陈述的事实感到苦恼的程度。你必须从每项提供的6个备选答案中选择其一来回答。

1. 感觉到你拥有的职权太小,无法完成分配给你的任务
   1. 从不苦恼　　4. 经常苦恼
   2. 很少苦恼　　5. 一直苦恼
   3. 有时苦恼　　6.(各选项)均不适用
2. 不清楚你的工作范围和职责是什么
   1. 从不苦恼　　4. 经常苦恼
   2. 很少苦恼　　5. 一直苦恼
   3. 有时苦恼　　6.(各选项)均不适用
3. 不知道有没有提升自己、发展自己的机会
   1. 从不苦恼　　4. 经常苦恼
   2. 很少苦恼　　5. 一直苦恼
   3. 有时苦恼　　6.(各选项)均不适用
4. 感觉有太重的工作负荷,不可能在一个正常的工作日之内完成任务
   1. 从不苦恼　　4. 经常苦恼
   2. 很少苦恼　　5. 一直苦恼
   3. 有时苦恼　　6.(各选项)均不适用
5. 无法满足不同上级相冲突的要求
   1. 从不苦恼　　4. 经常苦恼
   2. 很少苦恼　　5. 一直苦恼
   3. 有时苦恼　　6.(各选项)均不适用
6. 感觉还不足以胜任自己的工作
   1. 从不苦恼　　4. 经常苦恼
   2. 很少苦恼　　5. 一直苦恼
   3. 有时苦恼　　6.(各选项)均不适用
7. 不知道你的上司如何看待你以及如何评价你的表现
   1. 从不苦恼　　4. 经常苦恼
   2. 很少苦恼　　5. 一直苦恼
   3. 有时苦恼　　6.(各选项)均不适用
8. 无法获知完成你的工作所需要的信息
   1. 从不苦恼　　4. 经常苦恼
   2. 很少苦恼　　5. 一直苦恼
   3. 有时苦恼　　6.(各选项)均不适用
9. 不得不对你所认识的人做出影响其生活的决定
   1. 从不苦恼　　4. 经常苦恼

2. 很少苦恼　　5. 一直苦恼
3. 有时苦恼　　6. (各选项)均不适用

10. 觉得自己可能不为同事喜欢和接受
1. 从不苦恼　　4. 经常苦恼
2. 很少苦恼　　5. 一直苦恼
3. 有时苦恼　　6. (各选项)均不适用

11. 感觉无法改变顶头上司做出的将会影响到自己的决定和行动
1. 从不苦恼　　4. 经常苦恼
2. 很少苦恼　　5. 一直苦恼
3. 有时苦恼　　6. (各选项)均不适用

12. 不知道同事对自己有什么期望
1. 从不苦恼　　4. 经常苦恼
2. 很少苦恼　　5. 一直苦恼
3. 有时苦恼　　6. (各选项)均不适用

13. 觉得自己必须完成的工作的数量可能会妨碍工作完成的质量
1. 从不苦恼　　4. 经常苦恼
2. 很少苦恼　　5. 一直苦恼
3. 有时苦恼　　6. (各选项)均不适用

14. 感觉到你不得不在工作上做一些事与愿违的事情
1. 从不苦恼　　4. 经常苦恼
2. 很少苦恼　　5. 一直苦恼
3. 有时苦恼　　6. (各选项)均不适用

15. 感觉到你的工作会妨碍到你的家庭生活
1. 从不苦恼　　4. 经常苦恼
2. 很少苦恼　　5. 一直苦恼
3. 有时苦恼　　6. (各选项)均不适用

资料来源:于1964年经作者同意改编自:Robert L. Kahn & Donald M. Wolfe, *Organization Stress: Studies in Role Conflict and Ambiguity*, New York: John Wiley, 1964, pp. 424-425.

7.7 社　区

衡量社区的变量很有限。其中较早一个对于城市居民的公益心的测量的尝试是由E. L. Thorndike设计的。他的研究专论《我们的城市》(*Our City*, 1939)提供了评估美国城市人口质量的第一步认真尝试。美国310个人口超过30 000的城市的等级标准被确定下来。在Thorndike的著作《144个较小城市》(*144 Smaller Cities*, 1940)一书中,他为人口在20 000~30 000的城市划定了城市公益性的标准。这种方法所要求的

要素统计量不容易获得。Paul B. Gillen(1951),在《以职业分布作为城市衡量标准》(*The Distribution of Occupation as a City Yardstick*)一书中基于城市职业的分类提出了一种简捷的技巧。

过去的几十年表明,研究者越来越重视生活质量(或者说"什么是幸福")的研究。众多的研究者和研究组织正在探索在不同社区测量生活质量。其中,部分研究反映了具有吸引力的研究的重要性。可在这些研究中发挥作用的有关机构包括:当地商会、政府经济发展委员会、众多的其他相关政党。地方引以为骄傲的是促进了对社区成绩的测量。

社区生活质量的测量、比较评估和模式趋向,表现了应用社会学的先进之处。

参考文献

Gillen, Paul B. (1951). *The distribution of occupations as a city yardstick.* New York: Columbia University Press.

Thorndike, E. L. (1939). *Our city.* New York: Harcourt, Brace.

Thorndike, E. L. (1940). 144 *smaller cities.* New York: Harcourt, Brace.

研究应用

Haskitz, Alan. (1996). A community service program that can be validated. *Phi Delta Kappan*, 78 (2), 163-164.

LeSourd, Sandra J. (1997). Community service in a multicultural nation. *Theory into Practice*, 36 (3),157-163.

Loupe, Diane. (2000). Community service: Mandatory or voluntary? *School Administrator*, 57(7), 32-34, 36-39.

Morton, Keith, & Saltmarsh, John. (1997). Addams, Day, and Dewey: The emergence of community service in American culture. *Michigan Journal of Community Service Learning*, 4 (Fall), 137-149.

Neururer, Julie, & Rhoads, Robert A. (1998). Community service: Panacea, paradox, or potentiation. *Journal of College Student Development*, 39, 321-330.

Peterson, Elizabeth A. (1997). What can adults learn from community service? Lessons learned from Ameri Corps. *Community Education Journal*, 25(1-2), 45-46.

Skinner, Rebecca, & Chapman, Chris. (1999). Service-learning and community service in K-12 public schools. *Education Statistics Quarterly*, 1(4), 51-59.

White, Glen W. (1992). National and community service. *Journal of Disability Policy Studies*, 3 (2), 75-84.

Yates, Miranda, & Youniss, James. (1996). Community service and political-moral identity in adolescents. *Journal of Research on Adolescence*, 6, 271-284.

Yates, Miranda, & Youniss, James. (1997, April). *Engendering civic identity through community service.* Paper presented at the biennial meeting of the Society for Research on Child Development, Washington, DC.

7.7.1 社区态度量表

被测变量 态度的渐进程度在社区生活的这些领域比较明显:社区综合发展、生活条件、商业和工业、健康和娱乐、教育、宗教、青年文化、效用、通讯等。

描述 不同社区之间的交叉群体通过提交一系列陈述来定义"递进"的意义,这些陈述指出了递进和非递进的不同。这些陈述分为364个问题并以5分位的Likert量表来回答。代表性的领导集团独立判断每个问题是递进还是非递进。多方面的测试显示其中59个项目是很有识别力的。这59个项目被归纳为3个次量表,每个次量表有20个问题。这些量表被确定为:社区服务、社区整合、市政职责。

出处 Claud A. Bosworth的博士论文,所申请的学校为密歇根州大学,1954年。

信度 59个问题的量表,$r=0.56$。

效度 在采用递进方法和拒绝采用"递进"方法群体之间,总平均数在0.025水平上有显著差别。同时也发现,那些赞成量表所设计的测试人们对于社区发展另外阶段之态度选项的市民,也投票赞成下水道扩建计划。

效用 这个测量很容易掌握,适用于访谈和问卷调查。大约需要20分钟。

社区态度量表

	非常同意	同意	无所谓	不同意	非常不同意
社区服务量表					
1. 学校应该坚持学生的基本科学文化(3R)素质教育而忽略现今所注重的其他多数科目					
2. 多数社区已经非常好,即使它们没有开始新的社区发展计划					
3. 每一个社团应该更多地注重开展有益于发展人们的音乐与演讲口才的活动					
4. 以往的社区是更理想的居住环境					
5. 长期的发展比短期的受益更重要					
6. 在这个社区中,我们有很多对社区有益的组织					
7. 家庭和教堂承担了为年轻人步入婚姻和扮演父母角色做准备的职责					
8. 对于老年人的赡养义务应该归于他们自己和他们的家庭,而不是由社区来承担					
9. 社区中年轻人的活动节目太多了					
10. 在大多数社区里,学校已经很完善了					
11. 在社区事业的准备阶段往往花费了太多时间					
12. 成人教育应该成为地方学校规划的一个基本部分					

续表

	非常同意	同意	无所谓	不同意	非常不同意
13. 在社区中,人们的身体健康状况应该由医生来负责					
14. (社区某些成员的)精神疾病与全社区是没有关系的					
15. 一个现代的社区应该提供社会中介服务					
16. 教堂充分地满足了人们精神上的需求					
17. 为了发展得更好,社区应为人们提供另外的娱乐消遣设施					
18. 一般而言,教徒是较好的公民					
19. 人们的社交需求是他们自己以及他们家庭自己的事情,而不应当由社区来负责					
20. 教堂的拓展和分布应该与人口的增长相一致					
社区整合量表					
21. 如果一个社区没有制订长久的发展计划,那对于整个(发展)来说将是不利的					
22. 工业的发展应当包括重点扶持本地工业					
23. 对于每一个公民来说,首要的职责在于为自己的口袋多赚些钱					
24. 城镇中工业越多,人们的生活质量会越低					
25. 那些在社区发展中不积极参与的人会指责那些积极参与的人					
26. 有益于社区的就是有益于我的					
27. 每一个人都应当随他自己的意愿来处理他自己的事,而其他人的事应当由其他人自己处理					
28. 一个强有力的商会是有益于社区的					
29. 在社区中,商会的领导们漠视大多数居民福利					
30. 如果每一个人能各自做好自己的事,那么一个社区将会发展得更好					
31. 社区组织的成员可能只参加那些涉及他们自身利益的聚会					
32. 只有当社区作为一个整体进步的时候,我们的每一个人才能进步					
33. 不关注他所雇佣的人抱怨或意见的人不是好市民					
34. 如果我们使农民照料好农事,而我们照料好我们的事情,那情况会更好					

续表

	非常同意	同意	无所谓	不同意	非常不同意
35. 所有的联盟中到处都是共产党人					
36. 好市民支持所有新事物的广泛流传,包括那些跟他们以及他们组织不适应的新事物					
37. 好市民应当帮助少数群体解决他们的困难					
38. 在我们的社会中,农民有着突出的地位					
39. 市民应当只参加那些有助于发挥他们兴趣的组织					
40. 每个人都想着损人利己					
市政职责量表					
41. 忙碌的人们不该为城市规划负责					
42. 保持社区清洁的主要责任在于市政官员					
43. 在不增加税收的情况下,社区改进是好的					
44. 年轻人对我们的社区事务没有好评					
45. 一个进步的社区应能够提供充足的停车设施					
46. 政府官员应当在实施大型市政规划前关注一下民众情感					
47. 一个好市民应当乐于理解市政发展的领导					
48. 如果只是使少数人受扰,那就能相对而言较好的实现发展					
49. 社区发展应当只和社区的少数领导有关					
50. 如果更少的人在社区发展事业上花费时间的话,社区一定会发展得更好					
51. 那些很有时间的人应当有义务承担市政项目					
52. 社区的居住条件应当被改善					
53. 一个好市民应当为市政发展出谋划策					
54. 发展贫民区是浪费钱的举动					
55. 警察机关应当对外地人格外严格					
56. 大多数社区里的铺砌街道和道路都很好					
57. 即使需要增加税收,城市的污水系统也应当随着发展而扩充					
58. 有些人正计划到贫民区生活					
59. 我们面对的主要问题是高税收					

资料来源:量表由 Claud A. Bosworth 制作

7.7.2 社区服务行为的计分卡

被测变量 在社区服务中的个体参与。

描述 计分卡是一种估计社区服务中了解个人参与度的专用指数。列举了15种组成社区服务活动的可能的行为项目。用0~15分记录下来,对每一个参与的项目都进行加权。

出处 量表由 Delbert C. Miller 设计。

信度 未做过测试。

效度 停留在表面效度。

计分 基于中产阶级社区100个成年人随机抽样调查。他们的状况为下:10~15分:活跃的社区参与成员;6~9分:普通参与成员;0~5分:低参与成员。

效用 计分卡操作少于4分钟,适用于个人和群体评估。

社区服务行为计分卡

每一个回答"是"的计1分

财政捐资——在过去的一年中,你是否

——捐钱给社区竞选活动?

——捐钱给教堂?

——捐钱给其他慈善事业?

一般活动——在过去的一年中,你是否

——任职于任何市政规划职能部门?

——任职于任何从事提高市民生活的委员会?

——支持任何市政活动规划的领导?

社区论点与问题——在过去的一年中,你是否

——关心市政问题?

——与不至一个人经常讨论市政问题?

——劝说其他人坚持某一立场?

——从其他人那儿得到建议?

——与主要领导人谈论问题?

——参观社区组织或者董事会议以了解情况?

——写信、发送传单,或者举行家庭会议?

群体活动——在过去的一年中,你是否

——属于一个或者多个涉及社区问题的组织?

——组织团体访问,或者邀请社区官员加入你的组织?

——总分

7.8 社会参与

这一部分包括关于 Chapin 的社会参与量表的信息,这一量表是关于各种专业的、市民的或社会的非官方组织(等)的参与情况的一般性量表。这一量表一般在总体参与模式是一个重要变量的时候得到应用。闲暇参与和娱乐量表使研究者能够得到闲暇模式的详尽情况,并可以得到每一个被访者在参与和娱乐方面回答的得分。

7.8.1 Chapin 社会参与量表

被测变量 个人或者家庭在社区群体与机构中的参与程度。

描述 这是一个测量领导群体的 Guttman 类型量表,再现系数为 0.92 或者 0.97。高分(18 分或更高)代表了名义领导人的成就。5 个组成部分如下:①成员,②出席,③财政捐资,④委员会成员,⑤拥有的办公室。

这些组成部分测量不同的维度:"成员"部分测量参与程度,而另外 4 个测量参与强度。正式群体中的"拒绝—接受"用 1、4、5 三个部分来测量,相关度依次为:r_{14}为 0.53 ~ 0.58,r_{15}为 0.36 ~ 0.40,r_{45}为 0.36 ~ 0.40。社会参与用 2、3 两个部分来测量,相关度为:r_{23}为 0.80 ~ 0.89。这些部分之间的其他相关度依次为:r_{12} = 0.88,r_{13} = 0.89,r_{24} = 0.60,r_{34} = 0.40,r_{35} = 0.35,r_{45} = 0.50 ~ 0.58。

出处 Stuart Chapin 所著的 *Experimental Designs in Sociological Research* (New York: Harper, 1995),附录 B,pp. 275-278。

信度 r 为 0.89 ~ 0.95。

效度 与 Chapin 的社会地位量表的得分做相关分析,其相关系数 r 为 0.62 ~ 0.66;与收入等级,r = 0.52;与职业群体,r = 0.63;与受过正式教育的年数,r = 0.54;夫妻之间,r = 0.76。

标准分数 职业群体的平均得分如下:①自由职业者:20,②管理人员和私营业主:20,③牧师:16,④熟练工人:12,⑤半熟练工人:8,⑥非熟练工人:4。

效用 用一张纸来记录被访者所属每一个群体的条目,五个栏目下共有五个条目,分别对应被访者回答的问题。将被访者的回答填入表格大概需要 10 ~ 15 分钟。这一量表也可以自己操作。

说明

1)以表格顶部 1 ~ 5 号五种参与类型所表明的夫妻目前所属群体的名字排列。不必填入该人成为组织成员的日期。如果该人的成员资格是属于一个完全地方性的群体,则必须填入 L,如果这一成员资格是属于某一州或全国性组织的地方分支,则填入 N。

社会参与量表

地址________ 编号________

丈夫________

年龄________ 教育程度________ 种族或国籍________

职业________ 收入________

组织名字	1. 成员	2. 出席	3. 财政捐资	4. 是否委员会成员(非名义上的)	5. 是否拥有办公室
1.					
2.					
3.					
4.					
5.					
6.					
7.					
8.					
9.					
10.					

总计________

日期________ 调查员________

资料来源:明尼苏达大学出版社,Minneapolis. 版权:1938 明尼苏达大学

2)一个组织意味着某种现存的和有组织的群体,虽未必一定,但是通常是在一个社区或者居民区的附近,比如俱乐部、集会处、组织(这种组织是商业的,或政治性的,或自由职业的,或宗教的)和工会等等;如果一个教会或者其他机构的亚群体以某种或多或少独立的分支来组织的,那么它们也被包括在内。

3)在"出席"这一项下只记录"出席"或者"不出席"的事实,而不考虑出席的会议数量(人们并未发现改变捐资数量对最终分数有足够影响)。

4)在"财政捐资"下只记录财政捐资或者没有捐资的事实,而不是捐资的数量(人们并未发现改变捐资数量对最终分数有足够影响)。

5)以前的成员资格、委员会的工作、拥有的办公室等不应被计算或记录从而用于计算最终得分。

6)最终得分的计算将成员资格计为1,出席计为2,捐资计为3,委员会成员资格计为4,拥有办公室计为5。如果一家之中是夫妻二人,则把他们的总分相加然后平均。结果就是这个家庭社会参与的平均得分。如果一家之中只有寡妇或者鳏夫,那么个人的参与得分就是这个家庭的得分。

7.8.2 闲暇参与和娱乐

被测变量 闲暇时间娱乐的方法和程度。

描述 该量表包括47个选项,都是被调查者可能参加的活动,每个选项的回答分五个量度,闲暇参与根据参与频率分为五个量度:1. 从不,2. 极少,3. 偶尔,4. 时常,5. 频繁;闲暇娱乐根据好恶程度分为五个量度:1. 非常不喜欢,2. 不喜欢,3. 无所谓,4. 喜欢,5. 非常喜欢。量表中的每个选项的回答是重复相同的。人们从不参加的那些项目就不给出喜欢—不喜欢的量度。

出处 C. R. Pace《他们去上学:对951位以前的大学生的研究》,(*They Went to Colleage*: *A Study of* 951 *Former University Students*, Minneapolis:明尼苏达大学,1941)

信度 还未知。

效度 闲暇参与,和收入的相关度为 $r=0.019$,和社会公民活动的相关度为 $r=0.40$,和文化程度的相关度为 $r=0.039$。

标准分数 下表是明尼苏达大学的研究中,问卷调查所获数据的概要:

	1924－1925		1928－1929	
	毕业生	非毕业生	毕业生	非毕业生
男性闲暇参与的中位数	125.00	123.24	132.29	131.72
男性闲暇乐趣的中位数	169.83	167.53	171.67	170.65
女性闲暇参与的中位数	137.90	137.50	133.97	
男性闲暇乐趣的中位数	177.73	178.75	180.38	176.87

效用 这个量表易于操作,而且可以个人自己操作,也易计分,省时省力。闲暇参与和乐趣的分数能同时获得,数据还能相互比较。

闲暇活动调查表

闲暇时间的利用日益被认为是个重要的社会问题,我们想知道人们是如何度过他们的闲暇时间的,这里有一个活动列表,页面左边的1、2、3、4、5五个数字指的是你做这些事情的频率,表格左上方是对这个几个数字的解释;页面右边的1、2、3、4、5五个数字指的是你喜欢这个活动的程度,右上方是对这几个数字的解释。如果你从未做过被提及的活动,则在左边的数字中选择1,以表示从未参与过,而右边的数字则不必选择,尽量不要漏答。

参与频率		是否喜欢
1. 从不 2. 极少 3. 偶尔 4. 时常 5. 频繁		1. 非常不喜欢 2. 不喜欢 3. 无所谓 4. 喜欢 5. 非常喜欢
1 2 3 4 5	1. 业余演剧	1 2 3 4 5
1 2 3 4 5	2. 公园、礼堂里的娱乐活动	1 2 3 4 5
1 2 3 4 5	3. 艺术工作(个人的)	1 2 3 4 5
1 2 3 4 5	4. 参加大的集体活动(如舞会、桥牌)	1 2 3 4 5
1 2 3 4 5	5. 参加小的宴会(聚餐,等等)	1 2 3 4 5
1 2 3 4 5	6. 以读书为乐	1 2 3 4 5
1 2 3 4 5	7. (政治、宗教)集会	1 2 3 4 5
1 2 3 4 5	8. 家庭闲谈	1 2 3 4 5
1 2 3 4 5	9. 玩牌	1 2 3 4 5
1 2 3 4 5	10. 教堂和相关群体	1 2 3 4 5
1 2 3 4 5	11. 跳舞	1 2 3 4 5
1 2 3 4 5	12. 约会	1 2 3 4 5
1 2 3 4 5	13. 在家中招待朋友	1 2 3 4 5
1 2 3 4 5	14. 看博览会	1 2 3 4 5
1 2 3 4 5	15. 和朋友小聚	1 2 3 4 5
1 2 3 4 5	16. 非正式会谈,如闲聊	1 2 3 4 5
1 2 3 4 5	17. 室内小群体消遣或运动,如篮球、排球	1 2 3 4 5
1 2 3 4 5	18. 室内单人消遣或运动——保龄球、健身、落袋弹子球、台球、手球	1 2 3 4 5
1 2 3 4 5	19. 针织、缝补、勾织等	1 2 3 4 5
1 2 3 4 5	20. 讲座(非课程)	1 2 3 4 5
1 2 3 4 5	21. 听广播、看电视	1 2 3 4 5
1 2 3 4 5	22. 文学写作、如诗、散文、故事等等	1 2 3 4 5
1 2 3 4 5	23. 看杂志	1 2 3 4 5

1	2	3	4	5	24. 看电影	1	2	3	4	5
1	2	3	4	5	25. 读报	1	2	3	4	5
1	2	3	4	5	26. 在家做家务	1	2	3	4	5
1	2	3	4	5	27. 作为成员参加组织或俱乐部聚会	1	2	3	4	5
1	2	3	4	5	28. 作为发起人参加组织或俱乐部聚会	1	2	3	4	5
1	2	3	4	5	29. 室外个人运动:高尔夫、骑马、滑冰、远足、网球	1	2	3	4	5
1	2	3	4	5	30. 室外集体运动:曲棍球、篮球等	1	2	3	4	5
1	2	3	4	5	31. 野餐	1	2	3	4	5
1	2	3	4	5	32. 玩乐器、唱歌	1	2	3	4	5
1	2	3	4	5	33. 购物	1	2	3	4	5
1	2	3	4	5	34. 呆坐、沉思	1	2	3	4	5
1	2	3	4	5	35. 观看体育比赛	1	2	3	4	5
1	2	3	4	5	36. 听交响乐或音乐会	1	2	3	4	5
1	2	3	4	5	37. 煲电话粥	1	2	3	4	5
1	2	3	4	5	38. 排演话剧	1	2	3	4	5
1	2	3	4	5	39. 外出旅游、参观	1	2	3	4	5
1	2	3	4	5	40. 去公共图书馆	1	2	3	4	5
1	2	3	4	5	41. 参观博物馆、艺术长廊	1	2	3	4	5
1	2	3	4	5	42. 义务工作——社区服务等	1	2	3	4	5
1	2	3	4	5	43. 写信	1	2	3	4	5
1	2	3	4	5	44. 特殊爱好——集邮、摄影、手工工艺、园艺和其他以上未包括进去的爱好	1	2	3	4	5
1	2	3	4	5	45. 钓鱼、打猎	1	2	3	4	5
1	2	3	4	5	46. 野营	1	2	3	4	5
1	2	3	4	5	47. 冲照片、画画	1	2	3	4	5

资料来源:引字 C. R. Pace 的著作《他们去上学:对 951 位以前的大学生的研究》,(Minneapolis:明尼苏达大学出版)版权:1941 年明尼苏达大学,允许使用。

7.9 工作组织中的领导

这部分包含两种可能会在工作组织中广泛应用的领导能力量表:第一种是领导能力评价量表,用以回答"作为一名主管你该如何做"的问题;第二种是管理行为描述量表,用以回答"你的主管实际上是怎么做的"问题。值得关注的是,这两种量表使得在同一个组织中对两种不同层次的领导能力进行测量成为可能。管理者与他(或她)顶头上司之间的关系,作为一种非常重要的关系早已显现出其重要性。这两种调查量表的应用,使得对这两种不同领导层的管理能力的比较成为可能。

上述两个量表,每个都是为特定的目的而设计,使用中也要服务于这种特定的目的。在测量一个领导者的个人自我定位时,可以应用领导能力评价量表。在获取直接下属对其管理者的感觉时,则可以使用管理行为描述量表。它可以发给雇员、管理者或者经理层人员填答。从日臻完善的量表结构来看,这两个量表经过反复的修改,已具有相当高的信度和效度。

一些有关组织绩效的测量或许也可以包括在其中。但是考虑到篇幅限制,这里没有加以评述。不过下述几种测量可以作为一种注释供组织研究者参考。

行政职位描述量表包括 191 个项目。这些都是工商业界行政职位的基本资质要求。该量表的第一部分是职位行为,第二部分是职位责任,第三部分是职位要求和限制,第四部分是职位特质(Hemphill,1960)。

责任、权威和授权量表用来测量被调查者对管理监督者责任、权威和授权意识的不同程度感知。

多重关系社会调查测量工作活动环境的人际变量。其中含有五个维度:人际关系描述层面、人际感知层面、人际关系现状层面、主观意愿层面以及人际交往消极层面(Tannenbaum, Weschler 和 Massarik,1961)。

在工作接触中使用社会计量分析是一种分析复杂组织结构的方法。这种方法可以让研究者通过描述联系人的活跃行为和群体之间的广泛接触,勾勒出结构协调的组织是一种什么模样。

参考文献

Hemphill, John K. (1960). *Dimensions of executive positions.* Columbus: Ohio State University, Bureau of Business Research.

Stogdill, Ralph M., & Shartle, Carroll L. (1955). *Methods in the study of administrative leadership.* Columbus: Ohio State University, Bureau of Business Research.

Tannenbaum, Robert, Weschler, Irving W., & Massarik, Fred. (1961). *Leadership and organization: A behavioral science approach.* New York: McGraw-Hill.

Weiss, Robert S., & Jacobson, Eugene. (1955). A method for the analysis of the structure of complex organizations. *American Sociological Review*, 20, 661-668.

7.9.1　领导能力评价表

被测变量　这一调查表用以测量领导者在结构和关心两个要素上的取向。结构反映的是个体构建和确立自己角色与其下属目标达成相协调一致的程度。在这个维度上的高分获得者,通过规划、交流信息、编制进度表、发掘新创意等诸如此类的方法,在指导团体行动中扮演着非常积极的角色。关心反映的是个体通过相互信任、尊重下属意见、重视下属情感,构建良好工作关系的适当程度。在这个维度上的高分获得者,表现为拥有和谐的、双向交流的人际关系氛围。低分获得者,则表明该组织中,领导者与团队成员之间更多的是非情感性关系。

描述　这个有 40 个项目的调查表被分为结构和关心两大部分。每个部分有 20

个项目。这些项目有从0到4的五个连续取值。得分情况依项目在整个维度中的取向而定。

出处 可以从"Reid London House"网站找到相应资料,其网址为:www. reidlondonhouse. com/reidlondhouse/tests/loq. htm;电话:1-800-221-8378 或 847-292-1900;传真:847-292-3402;e-mail:reidlondonhouse@ncs. com。版权归1960年 Science Research Associates, Inc. 这一量表第一次由 Ralph M. Stogdill 和 Alvin E. Coons 推出,在《领导者的行为:描述和测量》(*Leaderbahavor:It's Description and Mesurement*,1957)一书的121~133页中。

信度 关于31个领班3个月后的再测系数如下:$r = 0.80$(关心维度),$r = 0.74$(初始结构)。对24个空降兵军士的再测系数分别是:$r = 0.77$(关心维度),$r = 0.67$(初始结构)。折半信度在关心和初始结构上的估值分别是0.69和0.73。

效度 效度通过独立的领导能力测量之间的相互联系来评估。比如,主管者做的优点评价、同级评价,管理部门做的压力选择绩效报告,无领导小组状况测试。尽管这个量表可以测出一些有重要统计价值的相关关系,但是在采用特殊标准的情况下,效度相对较低。相关性和其他一些测量显示,领导能力评价问卷的得分高低与管理者的智力不相关。其他一些有效的领导能力态度调查表也没有显示高智力的管理者在领导能力评价问卷上获得高分的优势。

这个调查问卷所得分值能直接可靠地区分出不同职位的领导态度,同样还能评估领导能力训练的效果。

计分 这个评估工具提供了多种计分方案,比如手工计分副本模式或者"Quanta™ Software for Windows®"模式,还有光学扫描模式(optical scanning)。答题卡可以交给激光读卡器计分,也可以登陆或发邮件到 Reid London House 网站计分。

效用 这个量表在10~15分钟内就可较容易地完成,量表的分值也很容易计算。

参考文献

Fleishman, Edwin A. (1960). *A manual for administering the Leadership Opinion Questionnaire.* Chicago: Science Research Associates.

研究应用

Dapra, Richard A., Zarrillo, Deirdre L., Carlson, Thomas K., & Teevan, Richard C. (1985). Fear of failure and indices of leadership utilized in the training of ROTC cadets. *Psychological Reports*, 56(1), 27-30.

Dukes, David Jefferson. (1996). Brain topography of leadership: Neurophysiological correlates of the Leadership Opinion Questionnaire. *Dissertation Abstracts International*, 56, 7082B.

Fok, Lillian Y., Hartmann, Sandra J., Crow, Stephen M., & Moore, Alger. (1995). Use of the Leadership Opinion Questionnaire to predict managerial success in organizations: A longitudinal study. *Organization Development Journal*, 13 (1), 23-32.

Kaurala, Earl B. (1997). Leader attitudes and leader stress: Is there a connection? *Dissertation Abstracts International*, 57, 3765A.

Kuntonbutr, Chanongkorn. (1999). A comparative study between Thai and American subordinates' perception of managerial values in the banking industry. *Dissertation Abstracts International*, 59, 4473A.

Oakland, Thomas, Falkenberg, Bradd A., & Oakland, Christopher. (1996). Assessment of leadership in children, youth and adults. *Gifted Child Quarterly*, 40(3), 138-146.

Schippmann, Jeffery S., & Prien, Erich P. (1986). Individual difference correlates of two leadership styles. *Psychological Reports*, 59(2, Pt. 2), 817-818.

Tharenou, Phyllis, & Lyndon, John T. (1990). The effect of a supervisory development program on leadership style. *Journal of Business and Psychology*, 4, 365-373.

Thomas, Veronica G., & Littig, Lawrence W. (1985). A typology of leadership style: Examining gender and race effects. *Bulletin of the Psychonomic Society*, 23(2), 132-134.

Trewatha, Robert L., & Vaught, Bobby C. (1987). The role of preferred leader behavior, managerial demographics, and interpersonal skills in predicting leadership style. *Journal of Behavioral Economics*, 16(1), 99-107.

领导能力评价表

该调查表作为一个完整量表共含有40个项目。下表的例子是一些来自这个完整量表的项目。研究者可以根据这些项目来评估使用这个完整量表的可能性。

结构：

给工作组织中的成员布置特定的任务

1. 总是如此　2. 经常　3. 偶尔　4. 几乎不　5. 从不

施予竞争性工作团队的压力

1. 很大　2. 相当大　3. 有一定程度　4. 较小　5. 根本没有

批评平庸的工作表现

1. 总是如此　2. 经常　3. 偶尔　4. 几乎不　5. 从不

对按时完成任务的强调

1. 很多　2. 相当多　3. 有一定程度　4. 较弱　5. 根本没有

关心：

采纳工作组织成员的建议

1. 总是如此　2. 经常　3. 偶尔　4. 几乎不　5. 从不

帮助工作组织成员解决个人问题

1. 频繁　2. 比较频繁　3. 有时候　4. 偶尔　5. 几乎不

在解决重大问题前征得工作团队的赞同

1. 总是如此　2. 经常　3. 偶尔　4. 几乎不　5. 从不

资料来源：经作者 Edwin A. Fleishman 同意后《领导者评价表》(*Leadership Opinion Questionaire*)于1960年再版，版权归1960 Science Research Associates, Inc. 所有，Chicago, IL.

7.9.2 管理行为描述量表

被测变量 下属员工对领导能力行为的认识,来自于他们对其顶头上司的言行举止的感知。因素分析显示,初始结构和关心两条项目是区分领导绩效的最重要因素。初始结构反映了主管者推动团队成员相互协作完成任务的程度,关心反映的是管理者重视下属情感的程度。问卷中的这些问题可以用一个问题来概括:“你的主管实际上是怎么做的”。

描述 这份量表有48个关于领导能力的项目,被分成两个相互独立的部分:初始结构和关心。第一个部分包括了20个项目,第二个部分包括了28个项目。这些项目的值有从0到4五个连续分值。得分情况依项目在整个维度中的取向而定。关心维度的最高取值是112,初始结构的最大值是80。

出处 Edwian A. Fleishman,《工作组织领导行为描述》(A Leader Behavior Description for Industry),载《领导行为:描述与测量》(*Leader Behavior: Its Description and Measurement*, ed. Ralph M. Stogdilland AlvinE. Coons (Columbus: Ohio State University, Brureau of Business Research, 1957), pp. 103-119。

信度 足够多的样本显示,该问卷的再测信度系数取值范围是0.46~0.87。样本的折半信度显示,其取值范围是0.68~0.98。

效度 从对122个领班的调查结果来看,关心和初始结构之间的相关系数为-0.2。从对394名员工(对这122个领班进行描述的人)的调查来看,组间相关系数为-0.33。对176名空军士兵和(美国)后备军官训练队学员(描述了他们的长官实际行为)的调查显示,两个量表之间的相关系数是-0.5。这两个要素之间的非相关性可以得到证实。

领班的行为和作为独立指标的领班所属团队中的事故发生率、旷工、牢骚、工作完成量之间有相关关系。在生产部门,关心量表中的高分获得者,被领班的主管预设为低熟练级别和少旷工的员工;初始结构量表中的高分获得者,被预设为高熟练级别和经常旷工并对工作不满的员工。

标准分数 标准分值通过一些调查研究得到,其结果如下:

样本	维度			
	关心		初始结构	
	平均数	标准差	平均数	标准差
对122个领班的描述	79.8	14.5	41.5	7.6
对31个领班的描述	71.5	13.2	37.5	6.3
对31个领班的描述	73.0	12.7	40.7	7.3
对8个文职主管的描述	75.1	17.6	37.3	9.6
对60个普通领班的描述	82.3	15.5	51.5	8.8

效用 完成这份调查表可能需要10~15分钟。这份调查表在团队中的应用特别有效。这份问卷和领导行为描述量表并用,可能会得到两个可以相互比较的观点,即

一位主管思考自己应该如何做和这位主管的下属们对主管实际表现的评价。Fleishman 对此做了最好的研究总结(1957)。在 International Harvester 公司,Fleishman 做了很多这方面的研究。

参考文献

Fleishman, Edwin A. (1957). A leader behavior description for industry. In Ralph M. Stogdill & Alvin E. Coons (Eds.), *Leader behavior: Its description and measurement* (pp. 103-119). Columbus: Ohio State University, Bureau of Business Research.

研究应用

Asher-Shultz, Nancy R. (1990). A comparison between observer ratings of supervisory behavior and the perceptions of supervisors and subordinates. *Dissertation Abstracts International*, 51, 2603B.

Brodhead, Sheila A. (1992). The effect of role stress and social support on supervisory behavior. *Dissertation Abstracts International*, 52, 3934B.

Bryant, Scott E., & Gurman, Ernest B. (1996). Contingent supervisory behavior: A practical predictor of performance. *Group and Organization Management*, 21, 404-413.

Cherniss, Cary. (1988). Observed supervisory behavior and teacher burnout in special education. *Exceptional Children*, 54, 449-454.

Childers, Terry L., Dubinsky, Alan J., & Skinner, Steven J. (1990). Leadership substitutes as moderators of sales supervisory behavior. *Journal of Business Research*, 21, 363-382.

Giacalone, Robert A., & Pollard, Hinda G. (1990). Acceptance of managerial accounts for unethical supervisory behavior. *Journal of Social Psychology*, 130(1), 103-109.

Kennebrew, Johnny L., & Sistrunk, Walter E. (1989, November). *Principals' supervisory behavior and school climate.* Paper presented at the annual meeting of the Mid-South Educational Research Association, Little Rock, AR.

Kohli, Ajay K. (1989). Effects of supervisory behavior: The role of individual differences among salespeople. *Journal of Marketing*, 53(4), 40-50.

Maroldo, Georgette K. (1988). Private shyness, social loneliness, and supervisory behavior. *Organization Development Journal*, 6(3), 56-62.

Pullen, Deborah B., & Sistrnnk, Walter E. (1990, November). *Ideal and actual supervisory behavior as perceived by Mississippi Community College vocational instructors and their supervisors.* Paper presented at the annual conference of the Mid-South Educational Research Association, New Orleans, LA.

管理行为描述量表的修订表

项目序号	项　目
	关心:修订的关键指标
1.	当与别人产生分歧时,他拒绝让步。
2.	对他的下属领班给予人性化的帮助。
3.	当我们中有人工作完成得出色时,他会给予赞赏。

4. 他容易理解人。
5. 对我们的要求高过我们的能力。
6. 他对他的领班在私人问题上给予帮助。
7. 他当着别人的面批评他的领班。
8. 他支持他的领班,哪怕这使他自己不受欢迎。
9. 他坚持所有的事情必须按他的方法做。
10. 他赞同一个工作业绩好的领班应该得到好的报酬。
11. 他拒绝有关变革的建议。
12. 在改变其下属的责任范围时,他并没有首先和下属进行详细交谈。
13. 在对待下属上,他很少考虑到他们的情绪。
14. 他尽力使下属领班在更高的权威下保持好的声望。
15. 他拒绝改变行事的方式。
16. 他会"驾驭"犯错误的领班。
17. 他拒绝对他的行为给予解释。
18. 他在采取行动之前没有首先和领班商量。
19. 他强调下属的高昂士气很重要。
20. 他支持下属领班的行动。
21. 他对新思想的接受迟缓。
22. 他将所有下属领班放在与自己对等的地位。
23. 他的批评对事不对人。
24. 他希望变革。
25. 他让下属在与他交谈中轻松自在。
26. 他友善、容易接近。
27. 他推进下属领班的建议付诸实施。
28. 他在对重大事情采取措施前,征得下属领班的支持。

初始结构:修订的关键指标

1. 他鼓励加班。
2. 他尝试自己的新想法。
3. 他以铁腕手段管理。
4. 他批评平庸的工作。
5. 他讨论应该完成多少利润。
6. 他鼓励工作迟缓的领班作出更多的努力。
7. 他欢迎下属领班在自己之前实施新的想法。
8. 他给自己的下属布置特殊的任务。
9. 他要求下属领班能够为整个部门的利益作出牺牲。
10. 他坚持让下属领班在处理事情的每一个细节上符合标准规范。
11. 他一定注意到下属之间正逐渐形成各自的群体界限。

12. 他提供解决问题的新方案。
13. 他坚持下属领班在对问题作出决定之前要告诉他。
14. 他让别人以他们自认为最好的方式行事。
15. 他给竞争性工作团队以压力。
16. 他"刺激"下属领班做出更多努力。
17. 他对该做些什么、该如何去做,作出细节上的决定。
18. 他强调按时完成任务。
19. 他要求低效率的工作团队领班善于总结提高自己。
20. 他注重工作成果的数量。

注:大多数项目的答案有如下几种:1. 总是如此,2. 经常,3. 偶尔,4. 几乎不,5. 从不。

资料来源:Edwin A. Flishman(1957),"*A Leader Behavior Description for Industry*," in Ralph M. Stogdill & E. Coons (Eds.),Leader Behavior:Its Description and Measurement (Columbus:Ohio State Univerity,Bureau of Business Research),103-119. 使用获得授权。

士气和工作满意度 7.10

士气只是我们试图用以表示个人社会态度和心理结构的众多词汇之一。许多量表由那些能够表示不同态度的术语构成。Karl Schuessler 已经检验了许多量表及表中的项目。士气和其他术语似乎太有限,难以定义这块领域。Schuessler 采用社会生活感受术语。这个术语的内涵相当宽泛,足以将各项目的共同属性涵盖进去。在接下来的 Schuessler 的社会生活感受量表中,项目的相互关系将显示重要的维度。

Nancy Morse 与合作者建立了一组次级量表,测量内部工作满意度、成就自豪感、团队归属感,以及财政和工作地位。如果要用小型量表测量这些维度,研究者应该使用 Morse 量表。*

如果要用一份具有既定信度和效度的问卷项目获取更多信息,严谨的研究者应该查阅《雇员的动机和士气调查表》(*Questionaire Measures of Employee Motivation and Morale*,Patchen,1966)。这本专著通过分析工作动机,工作创新的兴趣,主动向主管表达不同看法,对工作环境改变的态度,以及辨别工作组织等项目,来评估该问卷的信度和效度。

* 对总体工作满意度量表批评性评论可参见 Robinson,Athanasiou,以及 Head 的论述(1969,pp. 99-103)

参考文献

Patchen, Martin, with Pelz, Donald C. , & Allen, Craig W. (1966). *Some questionnaire measures of employee motivation and morale: A report on their reliability and validity.* Ann Arbor: University of Michigan, Institute for Social Research.

Robinson, John P. , Athanasiou, Robert, & Head, Kendra B. (1969). *Measures of occupational attitudes and occupational characteristics.* Ann Arbor: University of Michigan, Institute for Social Research.

7.10.1 Schuessler的社会生活感受量表

被测变量 对自我决断的怀疑、对人们可信赖性的怀疑、情绪低落、工作满意度、对公众事务的热心、情绪高涨、人们的犬儒主义、政府的觉醒、对未来的展望、经济自主、情绪受挫,以及对事业的关注,这些社会生活感受项目选自最近50年美国100多份社会学量表中出现的100多个项目。新泽西普林斯顿"回答分析公司"从全国范围内随机抽取1 522个居民户中的成人作为受访对象。通过187个访谈员对这些受访者进行了237个项目的测试。

从现有各项目组中的最高相关强度可以发现,12份个人量表可以被识别。每一个项目可以代表一个维度,这些维度建立在要素项目间的高组间相关关系上,而这些要素项目则可以用来判别是否进行因素分析。

出处 Karl F. Schuessler,《社会生活感受测量》(*Measuring social life feeling*, San Francisco: Jossey-Bass,1982)。

信度 总体评估建立在六类项目的基础上:初始信度,Tucker-Lewis信度,回答缺省率,社会背景的关联,测量的简易性,代表者的数量(他喜欢的量表数量)。该评估认为8个量表的信度高,4个量表的信度一般。

效度 效度的构建有赖于仔细检验每个量表中的最好的三个项目,最接近作者意图的量表,最切合主题的量表和最优的次级量表,然后对或简单或复杂的解释进行Schuessler测量(用一个或两个词)。这样表面效度和准则效度都得到检验。

管理 量表可以在调查问卷中使用,也可以在面访中使用。这些问卷设计者与"回答分析公司"合作,并建立了三种方式表示现有量表项目。在这三种方式中,回答者被问及"你是否很赞同或很不赞同每一个陈述"。被研究者应用的这些量表有三种使用方式:①将这些陈述读给被调查者听;②将问卷交给受访者并简要说一下问卷的指导语;③交给被调查者一组卡片,每张卡片上有一个陈述,要求他们按照自己对这些陈述的赞同程度将每张卡片排序。

使用方法 每一种方法于每个项目可使用约500次,于每位被调查者可使用近67次。结果表明,大多数项目的回答模式和测试技术具有统计意义上的独立性。

计分 所有消极陈述的答案表示为"同意","经常",或者"有时"的计1分。积极陈述的答案为"不同意","很少",或者"从不"的赋值为2分。[1] 标准值、平均数、中位数和标准差由12份量表来检验。这些量表选自全美国1 522个成人被调查者样本(这些测量是由"回答分析公司"完成)。

效用 这些量表是能够测量这些所给变量的最简短量表。任何一个量表很容易在2~3分钟之内回答完。任何研究者可以利用最适合他(她)研究设计的量表。利用两个或更多这类量表,可以描绘出个人或团体的基本概况。这些概况可以反映相对的地位和财富。

参考文献

Schuessler, Karl F. (1982). How social background influences people's responses. In *Measuring social life feelings*. San Francisco: Jossey-Bass, pp. 73-78. Reprinted with permission of the author.

研究应用

Krebs, Dagmar, & Schuessler, Karl. (1989). Life feeling scales for use in German and American samples. *Social Indicators Research*, 21, 113-131.

Lucke, Joseph F., & Schuessler, Karl. (1987). Scaling social life feelings by factor analysis of binary variables. *Social Indicators Research*, 19, 403-428.

Newton, Rae R., Prensky, David, & Schuessler, Karl. (1982). Form effect in the measurement of feelng states. *Social Science Research*, 11, 301-317.

Reiser, Mark, Wallace, Michael, & Schuessler, Karl. (1986). Direction-of-wording effects in dichotomous social life feeling items. *Sociological Methodology*, 16, 1-25.

Schuessler, Karl. (1982). How social background influences people's responses. In *Measuring social life feelings* (pp. 73-87). San Francisco: Jossey-Bass.

社会生活感受量表1:对自我决断的怀疑

陈述:

1. 当你较起真儿来,这个世界几乎没有人能够让你去信任。
2. 生活中发生的事情很大程度上是一种偶然。
3. 如果机遇不好,我们将不可能取得成就。
4. 对发生在自己身上的事情无能为力。
5. 有时候,感到无法把握自己命运的方向。
6. 现如今,一个人必须尽情享受今天,明天的事可以听天由命。
7. 我遇到的麻烦比预想的还多。
8. 对我而言,天天都是老样子。
9. 这个世界复杂的让我难以理解。
10. 我后悔自己过去错过了很多机会。
11. 让小孩面对不确定的未来是一件不公平的事情。
12. 命运叵测,使人很难预订计划。
13. 如今,我很难对任何事情都乐观。
14. 赚钱没有正确错误之分,只有容易困难之别。

社会生活感受量表 2:对他人信任的怀疑

陈述:

1. 如今,我们很难想出谁值得真正相信。
2. 当较起真儿来,这个世界几乎没有人可以让你去信任。
3. 多数人都是可以信任的。
4. 陌生人一般可以值得信任。
5. 多数人对待别人诚实无欺。
6. 多数人不会真正关心别人过得如何。
7. 在我们这个社会,太多的人关心的只是自己,对别人缺少真正的关心。
8. 一些人对你友好只是为了从你这儿捞点好处。

社会生活感受量表 3:情绪低落

陈述:

1. 我感到很多事情与自己无关。
2. 即便是和朋友在一起,我多少也感受到一些疏远。
3. 我有时候感到被朋友们所遗忘。
4. 有时候我对自己感到很陌生。
5. 我就是情不自禁地认为自己的生命没有多大意义。
6. 很孤独或者疏远别人。[1]
7. 沮丧或很不开心。[1]
8. 感到无聊。[1]
9. 焦躁不安难以坐很长时间。[1]
10. 莫名其妙地感到隐约的不安。[1]

注 释

1. O =“经常”;S =“有时”

社会生活感受量表 4:工作满意度

陈述:

1. 我的工作的变化太少。
2. 我工作时容易产生厌倦。
3. 一定有更好的地方适合我去工作。
4. 我喜欢工作中享有更多的自由。
5. 我对影响自己工作的事情缺少决定权。
6. 除了钱以外,工作对我还有更重要的意义。
7. 我对自己所做的工作感到满意。
8. 我的工作给了我最好地发挥自己能力的机会。
9. 人们对自己所工作的地方有归属感。

社会生活感受量表5:热心于公共事业

陈述:

1. 公众对于政客们在自己的办公室做些什么无从监督。
2. 普通老百姓难以找到与公务员交流的渠道。
3. 普通民众对国家政策具有一定影响力。
4. 普通百姓对当地政府的治理有相当大的发言权。
5. 像我一样的老百姓喜欢对政府大加评论。
6. 普通老百姓对政府的决策有很大的影响力。
7. 政府一般对民众的反映比较敏感。
8. 我通常对当地的选举比较感兴趣。
9. 积极参加政府和社会事务的人能够控制大量社会事务。
10. 如果我们对每件事情都给予重视,这个世界将会更富有成效。

社会生活感受量表6:情绪高涨

陈述:

1. 我时常感到无法控制自己命运的方向。
2. 在制订计划时,多数情况我都确信自己可以完成它。
3. 孩童时代的我比现在的我更快乐。
4. 我已经非常幸福了。
5. 我从生活中找到了很多乐趣。
6. 我对自己努力工作得到的结果很满意。
7. 我的前景一片光明。
8. 随着年龄的增长,我的处境越来越好。
9. 我和大部分人都有共同利益。
10. 这段时间我好像在原地踏步,没有变化。
11. 我现在做的事情都有明确的目的性。

社会生活感受量表7:大众犬儒主义(people cynicism)

陈述:

1. 在一个社会中,如果人人都只为自己考虑,人们之间将很快失去相互信任。
2. 绝大部分人都知道如何安排自己的生活。
3. 我们这个社会太多的人都只是考虑自己,很少真正顾及他人。
4. 我们社会中的许多人都很孤独,与他们身边的人没有什么联系。
5. 一些人对你友好只是想从你这儿捞点好处。
6. 许多人不知道该如何安排自己的生活。

社会生活感受量表8:对政府的幻觉化的觉醒

陈述:

1. 大多数超市都是诚实经营。
2. 政府渐渐限制了我们的自由。
3. 大多数政客关心的只是他们自己的利益,而非公众利益。
4. 我对现在的政府毫无信心。
5. 政府被少数掌权的人所控制。
6. 给政府公务人员写信毫无用处,因为他们通常并不会真正关心普通民众的问题。
7. 公务员是为老百姓工作,并非为他们自己工作。
8. 我们地方政府乱花纳税人的钱。
9. 就我的观点来看,这个国家处于病态之中。

社会生活感受量表9:未来展望

陈述:

1. 政府逐渐限制了我们的自由。
2. 我对现在的政府毫无信心。
3. 我们父辈所坚持的许多事情已经不合时宜了。
4. 尽管事物在不断变化,但是人们还是能够根据现在情况对未来作出预见。
5. 许多普通老百姓的处境正变得更差,而不是更好。
6. 前景一片黯淡。
7. 在今后的几年中将会有更多的人失业。
8. 很容易交到朋友。
9. 国家的未来很难确定。
10. 我的前景显得非常光明。
11. 如果我们对每件事情都给予重视,这个世界将会更富有成效。
12. 就我的观点来看,这个国家处于病态之中。

社会生活感受量表10:经济自决

陈述:

1. 个人贫困是因为他们不够努力。
2. 任何一个人,只要他愿意为生活拼搏,他就会提高他的生活水平。
3. 绝大部分人有很大的自由去决定自己的命运。
4. 生活贫穷的人只要努力拼搏就能改善他们的生活。
5. 我们国家有许多穷人,他们几乎难以提高他们的生活水平。

社会生活感受量表11:受挫感

陈述:

1. 我对发生在自己身上的事情无能为力。
2. 我觉得自己的身体条件很不错。

3. 这个世界复杂得我难以理解。
4. 与别人相比,我的生活不太好。
5. 我发现如今很难对任何事情保持乐观态度。
6. 我最近好像原地踏步、毫无进步。
7. 我不能给别人太多帮助。
8. 处于社会的上层。
9. 对于刚完成的事情感到满意。

社会生活感受量表 12:对事业的关注

问题:在最近几周里,这些事情是否经常出现在你的头脑里?
1. 金钱。
2. 工作。
3. 婚姻。
4. 晋升。
5. 抚养小孩。
6. 未来。

注 释

1. 关于基本原理、二分位量表(同意/不同意),见 Schuessler(1982,pp. 11-41)。

态度、价值和道德量表 7.11

这一节包括涉及当前利益和事务的态度和价值量表。在 20 世纪 60 年代和 70 年代早期,随着反文化运动的冲击,(研究者)大量的注意力被吸引到“个人与社会异化”的研究上来。社会学者响应了这种需求。他们这样的研究兴趣提高了 Srole Anomia 量表、Dean 疏远量表,以及其他此类量表在当前的使用率。虽然对“个人与社会异化”的兴趣减弱了,但是象征无能为力形象的大政府、大企业、大宗教团体、大劳工组织,以及大城市留存下来,并决定着人们的命运。因为这个原因,Neal 和 Seaman 的“无力感”量表(powerlessness scale)被选中,作为主流社会结构中个人和团队疏远的代表维度之一。

为了从现有的纷繁芜杂的量表中寻找更多信息,研究者要检验职业态度和职业特征测量(1969)、政治态度测量(1969),以及 John P. Robinson 和他的合作者们做的社会心理态度测量的表格内容。

参考文献

Robinson, John P., Athanasiou, Robert, & Head, Kendra B. (1969). *Measures of occupational attitudes and occupational characteristics.* Ann Arbor: University of Michigan, Institute for Social Research.

Robinson, John P., Rusk, Jerrold G., & Head, Kendra B. (1969). *Measures of political attitudes.* Ann Arbor: University of Michigan, Institute for Social Research.

Robinson, John P., & Shaver, Phillip R. (1973). *Measures of social psychological attitudes* (Rev. ed.). Ann Arbor: University of Michigan, Institute for Social Research.

7.11.1 异化和失范(alienation and anomie)测量

Daniel Bell 提出的"异化再发现"鼓励了学者们建立更多量表来测量这一现象。研究表明一些独立因子将被加以区分。Melvin Seeman(1959)阐述了五重的分类:无力感(powerlessness)、无意义(meaninglessness)、无规范(normlessness)、孤立(isolation),以及自我疏远(self-estrangement)。这些因子加以区分并测量这些已建构的量表。第一要素是,无力感,由黑格尔、马克思和韦伯讨论的,关于工人和他们控制自身经济命运的分离,他们的无助感,以及他们被利用来满足不是他们本身目的的目的。韦伯认为在工业社会,科学家、公务员和教授也同样与自身工作的控制相分离(同样不能控制自己的工作)。

Dwight Dean 建立了三个次级量表测量无力感、无规范和社会孤立感。他将这三个次级量表合在一起组成异化量表。他认为,这种相互关联模式表明异化可以被看成复合型概念。但是,这些次级量表中显示的变量表现了足够的独立性,这也证明,它们可以被视为独立变量。Arthur G. Neal 和 Solomon Retting(1963),就曾利用因子分析发现了无力感、无规范和 Srole 的失范量表具有独立结构的经验性根据。如果要求最大限度的精确性,那么这些次级量表在测量中就能派上用场。现在使用的这些量表,异质性高同质性低。

希望参考 Srole 量表和 Dean 量表研究者可以参见 Srole(1956)和 Dean(1961)的著作。研究者也许希望检验一下其他一些旨在测量异化和失范的量表,包括 McClosky 和 Schaar 的"异化的心理维度"(McClosky & Schaar,1965),和 Nettler 的"异化的测量"(Nettler,1957)等量表。如果要全面的例举、评估这些量表,可以参见 John P. Robinson 和 Phillip R. Shaver (1973,esp. pp. 254-294)的有关作品。

参考文献

Dean, Dwight G. (1961). Alienation: Its meaning and measurement. *American Sociological Review*, 26, 753-758.

McClosky, Herbert, & Schaar, John H. (1965). Psychological dimensions of anomie. *American Sociological Review*, 30, 14-40.

Neal, Arthur G., & Rettig, Solomon. (1963). Dimensions of alienation among manual and nonmanual workers. *American Sociological Review*, 28, 599-608.

Nettler, Gwynn. (1957). A measure of alienation. *American Sociological Review*, 22,670-677.
Roberts, Allan H., & Rokeach, Milton. (1956). Anomie, authoritarianism, and prejudice: A replication. *American Journal of Sociology*, 61, 355-358.
Robinson, John P., & Shaver, Phillip R. (1973). *Measures of social psychological attitudes* (Rev. ed.). Ann Arbor: University of Michigan, Institute for Social Research.
Seeman, Melvin. (1959). On the meaning of alienation. *American Sociological Review*, 24,783-791.
Srole, Leo. (1956). Social integration and certain corollaries: An exploratory study. *American Sociological Review*, 21,709-716.

研究应用

Furst, Edward J. (1975). *Measuring human-relations attitudes and values with situational inventories.* Fayetteville: Arkansas University, College of Education.
Gable, Robert K. (1993). *Instrument development in the affective domain: Measuring attitudes and values in corporate and school settings.* Norwood, NJ: Kluwer Academic.
Hullett, Craig R., & Boster, Franklin J. (2001). Matching messages to the values underlying value-expressive and social-adjustive attitudes: Reconciling an old theory with a contemporary measurement approach. *Communication Monographs*, 68, 133-153.
Kerlinger, Fred N. (1972, April). *The study and measurement of values and attitudes.* Paper presented at the American Educational Research Association, Chicago, IL.
Larkin, Caroline M. (1971). A longitudinal analysis of changes in attitudes and values during the undergraduate years as measured by the college student questionnaires. *Dissertation Abstracts International*, 32, 1880A.
Linn, Lawrence S., DiMatteo, M. Robin, Cope, Dennis W., & Robbins, Alan. (1987). Measuring physicians' humanistic attitudes, values, and behaviors. *Medical Care*, 25, 504-515.
Penn, John R. (1973). Values and attitudes as measures of intergenerational differences. *Dissertation Abstracts International*, 33,4531A-4532A.
Schumacher, Joseph E. (1992). Decision error tolerance and attitudes toward crime control: Measuring values toward judicial decisions. *Dissertation Abstracts International*, 52,4961B-4962B.
Suich, Paul S. (1992). Social interest: Triangulating the construct with measures of rational attitudes, communal values, and cooperative behavior. *Dissertation Abstracts International*, 52, 4988B.
Watral, David M. (1979). Relationships among values structure, attitudes and measured personality. *Dissertation Abstracts International*, 39, 5599B.

婚姻和家庭 7.12

婚姻调适是家庭研究中广泛应用的一个概念。学者尽量从概念和方法论上着手提高测量水平。一个通行的工具就是由 Graham B. Spanier 和 Erik E. Filsinger 建立的二元调适量表(the dyadic adjustment scale,DAS)。它为伴侣关系,不管是已婚还是未婚,提供了一种标准化的评估。尽管这超出了家庭社会学的研究范畴,但二元调适量

表在调适家庭关系的广泛领域内仍富有利用价值。该量表设计精细,具有很高的信度和效度。

7.12.1 二元调适量表

被测变量 婚姻调适是一个过程。它的结果决定于下述一些要素达到的程度:①令人生厌的婚姻差异;②婚期紧张和个人焦虑;③婚姻满意度;④双方凝聚力;⑤在对婚姻发挥功能有重要意义事情上的一致性。二元调适量表的组成部分包括,双方满意度、双方凝聚力、双方一致性和情感表达。

描述 二元调适量表有32个项目,Likert式问卷基本上使用5-,6-,和7-分值的答案。二元调适量表抽取自300个项目的初始组合形式,这300个项目都在前面的婚姻调适量表中使用过。那些先前测量中被忽视的新项目加进了这次婚姻调适测量中,而那些重复项目被删除了。三个专家根据前面给出的婚姻调适的定义,检查了这些项目的内容效度。根据这些专家的一致意见,大量的项目被删除。这样,这份量表问卷,包括一些人口统计项目,共有225个项目组成。

这份问卷曾在宾夕法尼亚中部,调查218个已婚者目标样本中使用过。问卷还调查了一些已从宾夕法尼亚国民中心获得离婚判决的人。调查时间是在这些人收到离婚判决书前的12个月中。这些测试的内容体现在已婚和离婚夫妇平均得分上。分辨力低于0.01的项目被删除,40个保留的变量用于因子分析,最终共保留了32个项目。

因子分析产生了四个相互关联的维度:双方一致性(dyadic consensus,夫妻在影响双方关系的重要事情上达成一致的程度),双方凝聚力(dyadic cohesion,夫妻一起共同参与活动的程度),双方满意度(dyadic satisfaction,夫妻双方对目前夫妻关系的满意程度和对保持这种关系的约定),以及情感表达(affectionac expression,双方对夫妻关系中情感表达和性生活的满意程度)。二元调适量表中还同时包括一些次级量表。1 000多份发表过的研究成果表明用过二元调适量表。90%以上的此类研究的研究对象包括了已婚者。

出处 多元健康系统,908 Niagara Falls Blvd., North Tonawanda, NY ,14120-2060,电话:1-800-456-3003,网址:www.mhs.com.这一量表出现在Erik E. Filsinger(1983)所著的《婚姻和家庭评估:家庭治疗手册》(*Marriage and Family Assessment:A Source Book for Family Therapy*)第8章的"二元调适量表"(CA:Sage)。

信度 Spanier指出了一些反映内部一致性信度的系数。如双方调适系数值是0.96,双方一致性系数值是0.90,双方凝聚力系数值是0.86,双方满意度系数值是0.94,以及情感表达系数值是0.73。Cronbach提出整个二元调适量表的系数值是0.96。

Filsinger和Wilson(1983,1984)指出,丈夫和妻子的拟似系数(similar coefficient)分别是,双方调适系数值是0.94、0.93;双方一致性系数值是0.91、0.88;双方凝聚力系数值是0.85、0.80;双方满意度系数值是0.82、0.84;以及情感表达系数值是0.73、0.73。

效度 专家根据理论维度给出内容效度。这份量表区分了已婚者和离婚者样本,

以及痛苦者和无痛苦者样本。

二元调适量表符合理论结构,具有建构效度。二元调适量表和 Locke-Wallace 婚姻调适量表(Locke-Wallace Marital Adjustment Scale, Locke & Wallace, 1959)相关系数在已婚夫妇中达到0.86,在离婚的受访者中达到0.88。

效用 二元调适量表不仅用来测量个人婚姻调适,还用来研究夫妻对婚姻的调适。同时,这一量表还可以用来测量未婚男女双方调适程度。它可以用来诊断夫妻关系是否陷入苦痛状况,可以辨识夫妻关系中的隐性问题,可以通过比较原先得分和治疗后得分来评估夫妻关系治疗的效果。它还可以用来做长期跟踪调查。现在,二元调适量表已被翻译成好几种语言,被不同的民族和文化群体所使用。

管理 二元调适量表可以在任何时间送给被调查夫妇。不过,让他们在调查开始或预约时间前期填答问卷,问卷效果可能发挥得更为充分。被调查者应该独自完成问卷,在完成问卷前不可以相互讨论。直到他们检查问卷并选定量表中的各项分值以后,被调查者也不可以和治疗专家讨论。

计分 二元调适量表中的每项回答都赋予了指定数字作为分值。个人的得分就是各个项目的数字之和。对研究者和治疗学家而言,量表的整体得分情况是最有价值的指标。但是,检验次级量表和个人项目的回答也可以成为分析问题起因的线索。研究者提供的已婚和离婚样本的量表平均分值分别是114.8和70.7。

夫妻得分可以派生出几种情况——例如,通过增加个人得分来区别它们——但是,这种操作还没有经验或理论上的支持。就这一点而言,这样无助于理解夫妻得分分值。

参考文献

Filsinger, E. E., & Wilson, M. R. (1983). Social anxiety and marital adjustment. *Family Relations*, 32, 513-519.

Filsinger, E. E., & Wilson, M. R. (1984). Religiosity, socioeconomic rewards, and family development: Predictors of marital adjustment. *Journal of Marriage and the Family*, 46, 663-670.

Locke, Harvey J., & Wallace, Karl M. (1959). Short marital-adjustment and prediction tests: Their reliability and validity. *Marriage and Family Living*, 21, 251-255.

研究应用

Bernard, Vicki Renee. (1999). Perceptions of sex role egalitarianism and dyadic adjustment among African-American couples. *Dissertation Abstracts International*, 59, 4291A.

Casas, J. Manuel, & Ortiz, Silvia. (1985). Exploring the applicability of the Dyadic Adjustment Scale for assessing level of marital adjustment with Mexican Americans. *Journal of Marriage and the Family*, 47, 1023-1027.

Crane, D. Russell, Middleton, Kenneth C., & Bean, Roy A. (2000). Establishing criterion scores for the Kansas Marital Satisfaction Scale and the Revised Dyadic Adjustment Scale. *American Journal of Family Therapy*, 28(1), 53-60.

Deleonardo, Lisa A. (2000). An investigation of identity development, ego strength, and dyadic adjustment in lesbian women. *Dissertation Abstracts International*, 61, 1077B.

Fisiloglu, Huerol, & Demir, Ayhan. (2000). Applicability of the Dyadic Adjustment Scale for measurement of marital quality with Turkish couples. *European Journal of Psychological Assessment*, 16(3), 214-218.

Hackney, Harold, & Bernard, Janine M. (1990). Dyadic adjustment processes in divorce counseling. *Journal of Counseling and Development*, 69(2), 134-143.

Hunsley, John, Best, Marlene, Lefebvre, Monique, & Vito, Diana. (2001). The seven-item short form of the Dyadic Adjustment Scale: Further evidence for construct validity. *American Journal of Family Therapy*, 29(4), 325-335.

Lim, Ben K., & Ivey, David. (2000). The assessment of marital adjustment with Chinese populations: A study of the psychometric properties of the Dyadic Adjustment Scale. *Contemporary Family Therapy: An International Journal*, 22,453-465.

Prouty, Anne M., Markowski, Edward Mel, & Barnes, Howard L. (2000). Using the Dyadic Adjustment Scale in marital therapy: An exploratory study. *Family Journal: Counseling and Therapy for Couples and Families*, 8(3), 250-257.

Wilson, Stephan M., Larson, Jeffry H., McCulloch, B. Jan, & Stone, Katherine L. (1997). Dyadic adjustment: An ecosystemic examination. *American Journal of Family Therapy*, 25(4), 291-306.

7.12.2 PREPARE 详细目录量表

被测变量 在未婚夫妇关系中与个人的、人际间的,以及外部的争议相关的主要问题。

描述 PREPARE 是一种自填式问卷。它含有 165 个 Likert 式的项目,每个项目的答案赋有 5 个分值(从非常同意到非常不同意)。问卷的主要目的是为评估夫妻关系问题提供一种诊断工具。PREPARE 涵盖婚姻预期、人格问题、交流、冲突解决、财务管理、休闲活动、性期望、孩子与养育、家庭和亲友、角色关系、精神信仰、夫妻亲密程度、家庭紧密程度、夫妻关系的弹性,以及家庭关系的弹性。已订婚的夫妇可以借助 PREPARE 和 PREPARE-MC 辨识他们自己的独特性问题。这样他们可以对即将到来的婚姻生活作出切合实际的评估。

出处 PREPARE 的指南和材料取自于 Life Innovations 公司,地址 P. O. Box 190, Minneapolis, MN 55440-0190,网址 www. lifeinnovations. com.

信度 从全国范围内抽取的 5 178 个样本来看, Cronbach 的 α 系数取值范围是 0.49 ~0.88。

研究应用

Fowers, Blaine J. (1991). His and her marriage: A multivariate study of gender and marital satisfaction. *Sex Roles*, 24(3-4), 209-221.

Fowers, Blaine J., Montel, Kelly H., & Olson, David H. (1996). Predicting marital success for premarital couple types based on PREPARE. *Journal of Marital and Family Therapy*, 22(1), 103-119.

Fowers, Blaine J., & Olson, David H. (1992). Four types of premarital couples: An empirical typology based on PREPARE. *Journal of Family Psychology*, 6 (1), 10-21.

Greeff, Abraham P., & Malherbe, Hildegarde L. (2001). Intimacy and marital satisfaction in spouses. *Journal of Sex and Marital Therapy*, 27(3), 247-257.

Larsen, Andrea S., & Olson, David H. (1989). Predicting marital satisfaction using PREPARE: A replication study. *Journal of Marital and Family Therapy*, 15(3), 311-322.

Montel, Kelly Huott. (1996). Predicting marital status outcome and satisfaction based on an empirical typology of engaged couples. *Dissertation Abstracts International*, 56, 7097B.

Rosen Grandon, Jane R. (1998). The relationship between marital characteristics, marital interaction processes, and marital satisfaction. *Dissertation Abstracts International*, 59, 1792A.

7.12.3 生活改善量表

生活改善公司(www.lifeinnovation.com)是建立在明尼阿波利斯的调查咨询中心。该中心主要设计一些工具用以评估、诊断和治疗准备结婚和已婚夫妇的关系。一些受过良好教育的家庭和心理咨询师管理着这家富有创意的公司。除了 PREPARE 详细目录以外,该公司还创建了其他一些量表,在我们检验家庭动力时都值得关注。这些量表包括很多基于研究的工具,诸如夫妻量表、家庭量表和一套观察量表。所有这些量表从生活改善公司获得。

(1)提升夫妻满意度量表

David Olson 创建了 17 个项目的夫妻满意度量表。该量表有 10 个项目是涉及满意度,7 个项目描述理想主义状态。

在对夫妻婚姻生活 10 个方面的调查中,婚姻满意度量表对此提供了全面的测量。这 10 个方面的调查内容包括了提升夫妻满意度项目中的主要类型:交流、冲突解决、角色、家庭财产利益关系、休闲时间、性关系、养育、家庭、亲友,以及信仰。

理想偏差量表测量的是被调查者在回答问题过程中所表现的乐观主义、现实主义,或者悲观主义的程度。这一量表可以作为理解个人认知偏差的一个有效参考。未婚夫妇倾向于过于乐观地描绘他们的夫妻关系。不过,不幸福的夫妇倾向过于悲观,而已婚夫妻则更多地倾向现实。

量表信度 资料显示量表有很高的信度($r=0.86$)。

效度 外部观察者对每个单独项目的评估说明了其良好的内容效度。量表的因子分析也证明了量表良好的结构效度。

样本项

- 我对我们彼此间的相互交流感到不太愉快,与我交流的配偶不理解我。

(2)ENRICH 夫妻量表

ENRICH 夫妻量表由 David H. Olson,David G. Fournier 和 Joan M. Druckman 提出。它包括三个 10 项的次级量表。这些量表可以用于研究婚姻满意度、交流和冲突解决

等内容。另外,它还含有一个7项的理想偏差量表。

在对夫妻婚姻生活九个方面的调查中,婚姻满意度量表对此提供了一个全面的测量。这九个方面的调查内容包括了婚姻质量提高的主要类型:人格、角色任务、交流、冲突解决、家庭财产利益关系、休闲时间的支配、性关系、父母责任、与家庭及亲友的关系,以及信仰取向。

交流量表关注的是在与他人交流中,个人对彼此关系的个人感受、信条。量表的项目集中于考察交流双方享有共同重要情感和信念的满意程度,对配偶传递或接受信息方法的认知,以及受访者对配偶与其交流到何种程度的感受。

冲突解决量表用以评价个人的态度、感受以及对缓和与解决关系冲突的信心。量表中的项目集中于考察个体辨识和解决问题的开放视角,结束争议的策略和方法,以及他们对问题解决途径的满意程度。

理想偏差量表测量的是个人在回答这些问题时表现的乐观、现实,或悲观的程度。这个量表对理解个人的认知偏差具有参考价值。未婚夫妇倾向于过于乐观地描述他们的夫妻关系,而不幸福的夫妻对夫妻关系过于倾向悲观,已婚的夫妇则更多地倾向于现实。

量表信度 资料显示这些量表有很好的信度。婚姻满意度量表的信度是 $r=0.86$,交流量表的信度为 $r=0.82$,冲突解决量表的信度是 $r=0.84$。

效度 外部观察者对每个单独项目的评估说明了其良好的内容效度。量表中的因子分析也证明了量表良好的结构效度。一致性效度在比较 Locke-Wallace 的婚姻调适量表中也有所显示。

样本项

- 我对我们如何做决定以及解决冲突的方式感到非常满意。(婚姻满意度项目)
- 我可以对我的另一半表达真实的感受。(交流项目)
- 为了结束争论,我往往很快就屈服。(冲突解决项目)

(3)家庭适应和内聚量表(Ⅱ)

David H. Olson,Joyce Portner 和 Yoav Lavee 在家庭调查中使用了家庭适应和内聚量表。这份30个项目的自填式量表可以用来评价 Circumplex 模型的两个主要维度:家庭内聚性和家庭适应。该量表的设计有助于指导家庭跨越生命周期。家庭适应和内聚量表Ⅱ可以和父母—未成年子女沟通量表或婚姻调适量表一起用来测量 Circumplex 模型的三个维度。

量表信度 资料显示该量表有良好的信度。内聚性项目的信度是 $r=0.87$,适应项目的信度是 $r=0.78$。

效度 许多研究结果显示该量表有良好的结构效度和一致性效度。

样本项

- 和别的人家相比,自己家庭成员之间关系更亲近。(家庭内聚性项目)
- 我们家尝试着用不同的方法处理家庭问题。(家庭适应性项目)

(4)家庭适应和内聚量表(Ⅲ)

David H. Olson,Joyce Portner 和 Yoav Lavee 采用的家庭适应性和内聚量表可以在

家庭实际工作中推荐使用。这份30个项目的自填式量表可以用来评价Circumplex模型的两个主要维度:家庭内聚性和家庭适应性。该量表的设计有助于指导家庭跨越生命周期。家庭适应与内聚量表Ⅲ可以和父母—未成年子女沟通量表或婚姻调适量表一起用来测量Circumplex模型的三个维度。由于比家庭适应与内聚量表Ⅲ具有更高的信度和效度,家庭适应与内聚量表Ⅱ更受推崇。家庭适应与内聚量表Ⅲ在指导实际家庭工作中受到推崇,是因为它涵盖的项目相对较少。

量表信度 资料明显显示该量表有良好的信度。内聚性项目的信度是$r=0.77$,适应性项目的信度是$r=0.62$。

效度 许多研究结果显示该量表有良好的结构效度和一致性效度。

样本项

- 空闲的时候,家里面的人喜欢聚在一起。(家庭内聚性项目)
- 我们家变换着处理任务的方式。(家庭适应性项目)

(5)家庭满意度

量表信度 资料显示该量表有良好的信度。内聚性项目的信度是$r=0.85$,适应性项目的信度是$r=0.84$。

效度 外部观察者对每个单独项目的评估说明了其良好的内容效度。量表中的因子分析也证明了量表具有良好的结构效度。

样本项

- 你对自己与家人亲近程度满意度如何?(家庭内聚性项目)
- 你对自己在家里说话的自由性满意度如何?(家庭适应性项目)

(6)父母—未成年子女沟通

Howard Barnes和David H. Olson建立的父母—未成年子女沟通量表是一种含有20个项目的自填式量表。该量表可以用来评价未成年子女及其父母对他们彼此交流的认知。这份量表包括两个次级量表,一个是自由的家庭交流量表,另一个是家庭交流中的问题量表。这两个次级量表用以测量未成年子女和其父母交流的积极方面和消极方面。

量表信度 资料显示两个次级量表有良好的信度。自由家庭交流量表的信度是$r=0.87$,家庭交流问题量表的信度是$r=0.78$。该总量表的整体信度同样较高,$r=0.88$。

效度 外部观察者对每个单独项目的评估说明了其良好的内容效度。量表中的因子分析也证明了量表具有良好的结构效度。

样本项

- 有些时候,我不敢将我想的东西告诉妈妈。
- 我的爸爸永远都是好的听众。
- 我的孩子不用我问就能将他(她)的感受告诉我。

(7)家庭力量量表

由David H. Olson,A. S. Larson和Hamilton I. McCubbin建立的家庭力量量表,共有12个项目,可以从自豪和协调维度测量家庭力量。

量表信度 资料显示两个次级量表有良好的信度。关于自豪的量表信度是 $r=0.88$,关于协调(一致)性的量表信度是 $r=0.72$。量表的整体信度是 $r=0.83$。

样本项

- 我们家里有很多冲突。

(8)生活质量

量表信度 资料显示父母量表和未成年子女量表都有很高信度,分别是 $r=0.92$ 和 $r=0.86$。

效度 外部观察者对每个单独项目的评估说明了其良好的内容效度。量表中的因子分析也证明了量表具有良好的结构效度。

样本项

- 你对自己家庭的满意程度如何?
- 你对自己必须的开支满意程度如何?

(9)临床评估量表

David H. Olson 和 Elinor Killorin 设计的临床评估量表最初是给治疗学家使用的。它可以从家庭内聚性、家庭适应性和家庭沟通维度来评估家庭成员的行为。该量表那时能够评估处于 Circumplex 模型中的夫妻地位或家庭地位。该量表是一种观察量表,而不是自填式量表。

量表信度 翔实的材料表明,该量表的三个次级量表具有高信度。关于家庭内聚性的量表信度是 $r=0.95$,关于家庭适应性的量表信度是 $r=0.94$,关于家庭沟通的量表信度是 $r=0.97$。

内部评价信度 家庭内聚性($r=0.95$),家庭适应性($r=0.91$)和沟通($r=0.97$)三个次级量表显示了很高的内部评价信度。

效度 外部观察者对每个单独项目的评估说明了其良好的内容效度。量表中的因子分析也证明了量表具有良好的结构效度。

7.13 个性测试:明尼苏达多相人格测试量表

此处仅从成百上千的既有人格目录中选择了一项,即明尼苏达多相人格检查表(Minnesota Multiphasic Personality Inventory-2™),或者称为 MMPI。它可以被描述使用,但不能直接拷贝使用。其由一组量表组成共包括 550 条表述。量表的内容全面,结构严谨,很多研究者都将其作为研究人格的最好量表。资料中描述的研究应用就证明了它的功用。

被测变量 MMPI(明尼苏达多相人格测试量表)包括 8 个效度量表,10 个临床量表,15 个内容量表,27 个内容因素量表,20 个补充量表,5PSY-5 量表,31 个次级临床量表,5 个高级自我表达次级量表,以及各种专门或特定指标。

描述 这个 MMPI 的设计,最初是在单项测试中,用来记录在所有临床重要阶段

的人格测试分值。这个量表本身包含了550个正确或错误的陈述。这些陈述涵盖了从个人的身体条件到其社会道德观的众多主题。完成这一量表,被调查者,即该家庭排在第一位的人,需要对量表中的所有表述作出"正确"、"错误"或"说不清楚"的回答。

出处 当MMPI方法于1942年公布时,Strake R. Hathaway(哲学博士)和J. C. Mckinley(医学博士)正在明尼苏达医学院求学。当前测试的作者包括re Yossef S. Ben-Porath, James N. Butcher, W. Grant Dahlstrom, John R. Graham, and Auke Tellegen.

信度 信度系数的取值介于0.71到0.83之间。

效度 Hathaway和McKinley主张,该量表是否出色的标准应该是依照神经精神病学专家的临床诊断病例作出的效度预测,而不是统计测量的信度和效度。

参考文献

Gough, Harrison G. (1947). Simulated patterns on the Minnesota Multiphasic Personality Inventory. *Journal of Abnormal and Social Psychology*,42,215-225.

Hathaway, Starke R., & McKinley, J. Chamley. (1951). *Manualforthe Minnesota Multiphasic Personality Inventory* (Rev. ed.). New York: Psychological Corporation.

Weisgerber, Charles A. (1951). The predictive value of the Minnesota Multiphasic Personality Inventory with student nurses. *Journal of Social Psychology*, 33, 3-11.

研究应用及推荐读物

Archer, R. P., Maruish, M., Imhof, E. A., & Piotrowski, C. (1991). Psychological test usage with adolescent clients: 1990 survey findings. *Professional Psychology: Research and Practice*, 22, 247-252.

Bagby, R. Michael, et al. (1997). Does clinical training facilitate feigning schizophrenia on the MMPI-2? *Psychological Assessment*, 9(2), 106-112.

Ben-Porath, Y. S., & Butcher, J. N. (1989). Psychometric stability of rewritten MMPI items. *Journal of Personality Assessment*, 53, 645-653.

Butcher, J. N. (1987). Computerized clinical and personality assessment using the MMPI. In J. N. Butcher (Ed.), *Computerized psychological assessment: A practitioner's guide* (pp. 161-197). New York: Basic Books.

Butcher, J. N. (1990). *MMPI-2 in psychological treatment.* New York: Oxford University Press.

Butcher, James Neal. (Ed.). (2000). *Basic sources on the MMPI-2.* Minneapolis: University of Minnesota Press.

Butcher, J. N., Graham, J. R., Dahlstrom, W. G., & Bowman, E. (1990). The MMPI-2 with college students. *Journal of Personality Assessment*, 54, 1-15.

Butcher, J. N., & Owen, P. L. (1978). Objective personality inventories: Recent research and some contemporary issues. In B. Wolman (Ed.), *Clinical diagnoses of mental disorders* (pp. 475-546). New York: Plenum.

Caldwell, A. B. (1991). MMPI-2 content scales: What you say is what you get? *Contemporary Psychology*, 6, 560-561.

Cloak, Nancy L., Kirklen, Leonard E., Strozier, Anne L., & Reed, James R. (1997). Factor analysis of Minnesota Multiphasic Personality Inventory-1 (MMPI-1) validity scale items. *Measurement and Evaluation in Counseling and Development*, 30(1), 40-49.

Dana, Richard H. (1995). Culturally competent MMPI assessment of Hispanic populations. *Hispanic Journal of Behavioral Sciences*, 17, 305-319.

Friedman, A. R., Webb, J. T., & Lewak, R. (1989). *Psychological assessment with the MMPI.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Graham, J. R. (1990). *MMPI-2: Assessing personality in psychopathology.* NewYork: Oxford University Press.

Graham, J. R., Timbrook, R. E., Ben-Porath, Y. S., & Butcher, J. N. (in press). Code-type congruence between MMPI and MMPI-2: Separating fact from artifact. *Psychological Assessment: A Journal of Consulting and Clinical Psychology.*

Greene, R. L. (1991). *The MMPI-2-MMPI An interpretive manual.* Boston: Allyn & Bacon.

Hostetler, K., Ben-Porath, Y. S., Butcher, J. N., & Graham, J. R. (1989, April). *New MMPI-2 subscales.* Paper presented at the annual meeting of the Society for Personality Assessment, New York, NY.

Jackson, D. N. (1989). *Basic Personality Inventory.* Port Huron, MI: Sigma Assessment Systems.

Levitt, E. E. (1990). A structural analysis of the impact of MMPI-2 on MMPI- 1. *Journal of Personality Assessment*, 55, 562-577.

Long, K. A., Graham, J. R., & Timbrook, R. E. (1994). Socioeconomic status and MMPI-2 interpretation. *Measurement and Evaluation in Counseling and Development*, 27(3), 158-177.

Megargee, Edwin I. (2001). *Classifying criminal offenders with the MMPI-2: The Megargee system.* Minneapolis: University of Minnesota Press.

Nichols, D. S. (in press). New MMPI-2 content scales. *Journal of Personality Assessment.*

Pancoast, D. L., & Archer, R. P. (1989). Original adult MMPI norms in normal samples: A review with implications for future developments. *Journal of Personality Assessment*, 53,376-395.

Schlenger, W. E., & Kulka, R. A. (1987, August). *Performance of the Keane-Fairbank MMPI scale and other self-report measures in identifying posttraumatic stress disorder.* Paper presented at the 95th Annual Convention of the American Psychological Association, New York, NY.

Sturmer, Paul J., & Gerstein, Lawrence H. (1997). MMPI profiles of black Americans: Is there a bias? *Journal of Mental Health Counseling*, 19(2), 114-129.

Tellegen, A. M. (1988, August). *Derivation of uniform T-scores for the restandardized MMPI.* Symposium presentation at the 96th Annual Convention of the American Psychological Association, Atlanta, GA.

Tsai, David C., & Pike, Patricia L. (2000). Effects of acculturation on the MMPI-2 scores of Asian American students. *Journal of Personality Assessment*, 74(2), 216-230.

Wrobel, Nancy Howells, & Lachar, David. (1995). Racial differences in adolescent self-report: A comparative validity study using homogeneous MMPI content measures. *Psychological Assessment*, 7(2), 140-147.

BUROS 研究所及 BUROS 心理测量年鉴 7.14

寻找大量的关于测量的评论信息的最重要资源是隶属于 Nebraska 大学的 BUROS 心理测量研究所(www. unl. edu/buros/)。该研究所有一个心理测量方面的连续出版物——BUROS 心理测量年鉴(mental measurement yearbook, MMY)。该年鉴的第 14 辑,由 Barbara Plake 和 James C. Impara 编写,并于 2001 年由 BUROS 心理测量研究所出版。下面所罗列的是年鉴中有关测量的名称,它们被归入 19 个类别,并按照字母顺序排列。如果你对有些测量感兴趣的话,可以查阅相关年鉴。需要注意的是,许多测量,特别是人格与智力方面的测量,只应该由受过严格训练的专业人士使用,比如:持有许可证的临床社会工作者、学校心理咨询师、临床心理咨询师等。

7.14.1 Achievement

ACER Tests of Basic Skills—Blue Series, 12th MMY
ACER Tests of Basic Skills—Green Series, 12th MMY
Achievement Test for Accounting Graduates, 13th MMY
Adult Basic Learning Examination, Second Edition, 11th MMY
Aprenda: La Prueba de Logros en Español, 13th MMY
Aprenda: La Prueba de Logros en Español—Segunda edicion, 14th MMY
Assessment and Placement Services for Community Colleges, 11th MMY
Basic Achievement Skills Individual Screener, 9th MMY
Basic Educational Skills Test, 9th MMY
Basic Skills Inventory, 9th MMY
Brigance Diagnostic Comprehensive Inventory of Basic Skills, 9th MMY
BRIGANCE Diagnostic Comprehensive Inventory of Basic Skills, Revised, 14th MMY
Brigance Diagnostic Inventory of Basic Skills, 9th MMY
Brigance Diagnostic Inventory of Essential Skills, 9th MMY
BRIGANCE Diagnostic Life Skills Inventory, 14th MMY
Bristol Achievement Tests, 9th MMY
California Achievement Tests, Fifth Edition, 13th MMY
Canadian Achievement Survey Tests for Adults, 14th MMY
Canadian Achievement Tests, Second Edition, 13th MMY
Canadian Tests of Basic Skills, Forms 7 and 8, 11th MMY
Closed High School Placement Test, 14th MMY
College Basic Academic Subjects Examination, 11th MMY
College Board Scholastic Aptitude Test, 9th MMY

Collegiate Assessment of Academic Proficiency, 13th MMY
Comprehensive Testing Program III, 13th MMY
Comprehensive Tests of Basic Skills, Fourth Edition, 11th MMY
CTB Portfolio Assessment System, 13th MMY
Curriculum Frameworks Assessment System, 14th MMY
Developing Skills Checklist, 12th MMY
Diagnostic Achievement Battery, Second Edition, 12th MMY
Diagnostic Achievement Test for Adolescents, Second Edition, 13th MMY
Diagnostic Screening Test: Achievement, 9th MMY
Differential Ability Scales, 11th MMY
Distar Mastery Tests, 9th MMY
Early School Assessment, 12th MMY
End-of-Course Tests, 11th MMY
Evaluation of Basic Skills, 14th MMY
GOALS: A Performance-Based Measure of Achievement, 13th MMY
Guidance Centre Classroom Achievement Tests, 11th MMY
Hammill Multiability Achievement Test, 14th MMY
High School Placement Test—Open Edition, 14th MMY
High-School Subject Tests, 11th MMY
Individualized Criterion Referenced Testing, 12th MMY
Integrated Assessment System, 13th MMY
Iowa Tests of Basic Skills, Forms K, L, and M, 13th MMY
Iowa Tests of Educational Development, Forms K, L, and M, 10th MMY
Kaufman Functional Academic Skills Test, 13th MMY
Kaufman Survey of Early Academic and Language Skills, 12th MMY
Kaufman Test of Educational Achievement, 10th MMY
Kaufman Test of Educational Achievement [1998 Normative Update], 14th MMY
Kuhlmann-Anderson Tests, Eighth Edition, 9th MMY
Life Skills, Forms 1 and 2, 9th MMY
Maculaitis Assessment Program, Commercial Edition, 10th MMY
Metropolitan Achievement Tests, Seventh Edition, 12th MMY
Metropolitan Performance Assessment: Integrated Performance Tasks, 14th MMY
Monitoring Basic Skills Progress—Second Edition, 12th MMY, 14th MMY
Multilevel Academic Survey Test, 10th MMY
National Achievement Test [Second Edition], 11th MMY
National Educational Development Tests, 11th MMY
National Proficiency Survey Series, 11th MMY
Norris Educational Achievement Test, 12th MMY

Peabody Individual Achievement Test—Revised, 11th MMY
Peabody Individual Achievement Test—Revised [1998 Normative Update], 14th MMY
PLAN, 12th MMY
QUEST, 12th MMY
QUIC Tests, 11th MMY
Quick Score Achievement Test, 10th MMY
Richmond Tests of Basic Skills: Edition 2, 11th MMY
Riverside Curriculum Assessment System, 13th MMY
SAT. *See* College Board Scholastic Aptitude Test
Scaled Curriculum Achievement Levels Test, 12th MMY
Scholastic Aptitude Scale, 11th MMY
Scholastic Aptitude Test. *See* College Board Scholastic Aptitude Test
Spanish Assessment of Basic Education, Second Edition, 13th MMY
SRA Achievement Series, Forms I and 2, and Survey of Basic Skills, Forms P and Q, 11th MMY
Stanford Achievement Test, Ninth Edition, 13th MMY
Stanford Achievement Test—Abbreviated Version—8th Edition, 12th MMY
Stanford Early School Achievement Test, Third Edition, 11th MMY
TerraNova, 14th MMY
Test of Academic Achievement Skills—Reading, Arithmetic, Spelling, and Listening Comprehension, 13th MMY
Test of Academic Performance, 13th MMY
Tests of Achievement and Proficiency, Forms K, L, and M, 14th MMY
Tests of Adult Basic Education, Forms 7 & 8, 13th MMY
Tests of Adult Basic Education Work-Related Foundation Skills, 13th MMY
Tests of General Educational Development [The GED Tests], 11th MMY
Wechsler Individual Achievement Test, 13th MMY
Wide Range Achievement Test 3, 12th MMY
Wonderlic Basic Skills Test, 13th MMY
Woodcock-McGrew-Werder Mini-Battery of Achievement, 13th MMY

7.14.2 Behavior Assessment

AAMR Adaptive Behavior Scale—Residential and Community, Second Edition, 13th MMY
AAMR Adaptive Behavior Scale—School, Second Edition, 13th MMY
Aberrant Behavior Checklist, 12th MMY
Achieving Behavioral Competencies, 13th MMY

Adaptive Behavior Evaluation Scale, 12th MMY
Adaptive Behavior Evaluation Scale, Revised, 14th MMY
Adaptive Behavior Inventory, 10th MMY
Adaptive Behavior Inventory for Children, 9th MMY
ADD-H: Comprehensive Teacher's Rating Scale, Second Edition, 12th MMY
AD/HD Comprehensive Teacher's Rating Scale, Second Edition [1998 Revision], 14th MMY
Adjustment Scales for Children and Adolescents, 14th MMY
Adult Attention Deficit Disorder Behavior Rating Scale, 13th MMY
Adult Attention Deficit Disorders Evaluation Scale, 14th MMY
Assessment of Adaptive Areas, 14th MMY
Attention Deficit Disorder Behavior Rating Scales, 13th MMY
Attention Deficit Disorders Evaluation Scale, Second Edition, 12th MMY, 14th MMY
Attention Deficit Disorders Evaluation Scale Secondary-Age Student (The), 14th MMY
Attention-Deficit/Hyperactivity Disorder Test, 14th MMY
Attention-Deficit Scales for Adults, 14th MMY
BASC Monitor for ADHD, 14th MMY
Bay Area Functional Performance Evaluation, Second Edition, 12th MMY
Behavior Analysis Forms for Clinical Intervention, 9th MMY
Behavior Assessment System for Children, 13th MMY
Behavior Assessment System for Children [Revised], 14th MMY
Behavior Dimensions Rating Scale, 11th MMY
Behavior Dimensions Scale, 14th MMY
Behavior Disorders Identification Scale, 12th MMY
Behavior Evaluation Scale—2, 12th MMY
Behavior Rating Profile, Second Edition, 12th MMY
Behavioral and Emotional Rating Scale: A Strength-Based Approach to Assessment, 14th MMY
Brown Attention-Deficit Disorder Scales, 14th MMY
Burks' Behavior Rating Scales, 11th MMY
Camelot Behavioral Checklist, 9th MMY
Caregiver-Teacher Report Form, 14th MMY
Cayuga-Onondaga Assessment for Children with Handicaps. *See* C. O. A. C. H.: Cayuga-Onondaga Assessment for Children with Handicaps
Checklist of Adaptive Living Skills, 12th MMY
Child "At Risk" for Drug Abuse Rating Scale (The), 12th MMY
Child Behavior Checklist, 13th MMY

Child Sexual Behavior Inventory, 14th MMY
Children's Attention and Adjustment Survey, 12th MMY
C. O. A. C. H. : Cayuga-Onondaga Assessment for Children with Handicaps, Version 7.0, 11th MMY
Comprehensive Behavior Rating Scale for Children, 13th MMY
Conners' Continuous Performance Test, 14th MMY
Conners' Rating Scales, 11th MMY
Conners' Rating Scales—Revised, 14th MMY
Continuing Education Assessment Inventory, 9th MMY
Cooper-Farran Behavioral Rating Scales, 12th MMY
Developmental Teaching Objectives and Rating Form—Revised, 14th MMY
Devereux Behavior Rating Scale—School Form, 13th MMY
Differential Test of Conduct and Emotional Problems, 12th MMY
Dyadic Parent-Child Interaction Coding System: A Manual, 9th MMY
Early Childhood Attention Deficit Disorders Evaluation Scale (The), 14th MMY
Early Childhood Behavior Scale (The), 13th MMY
Early Screening Project, 14th MMY
Eby Gifted Behavior Index, 12th MMY
Emotional and Behavior Problem Scale, 12th MMY
Emotional or Behavior Disorder Scale, 13th MMY
Eyberg Child Behavior Inventory, 9th MMY
Forms for Behavior Analysis with Children, 11th MMY
Functional Needs Assessment, 12th MMY
Gilliam Autism Rating Scale, 13th MMY
Gordon Diagnostic System (The), 13th MMY
Guide to the Assessment of Test Session Behavior for the WISC-III and the WIAT, 13th MMY
Hahnemann High School Behavior Rating Scale, 9th MMY
Hutchins Behavior Inventory, 12th MMY
Independent Living Scales, 14th MMY
Instrument for Disability Screening [Developmental Edition], 9th MMY
Integrated Visual and Auditory Continuous Performance Test, 14th MMY
Interpersonal Trust Surveys, 14th MMY
Iowa Social Competency Scales, 10th MMY
Leadership Practices Inventory [Pfeiffer & Company International Publishers], 12th MMY
Leadership Practices Inventory—Individual Contributor [Second Edition], 14th MMY
Learned Behaviors Profile—Self/Other, 11th MMY

Learning Skills Profile, 13th MMY
Louisville Behavior Checklist, 10th MMY
Millon Behavioral Health Inventory, 9th MMY
Mother/Infant Communication Screening, 12th MMY
Motivation Assessment Scale, 14th MMY
Parent Behavior Checklist, 13th MMY
Parent Behavior Form, 11th MMY
Portland Problem Behavior Checklist—Revised, 14th MMY
Pre-Referral Intervention Manual[Revised and Updated Second Edition], 13th MMY
Preschool and Kindergarten Behavior Scales, 13th MMY
Preschool Behavior Rating Scale, 10th MMY
Pyramid Scales (The), 10th MMY
REHAB: Rehabilitation Evaluation (Hall and Baker), 14th MMY
Responsibility and Independence Scale for Adolescents, 12th MMY
Revised Behavior Problem Checklist, 11th MMY
Scales of Independent Behavior—Revised, 14th MMY
School Function Assessment, 14th MMY
School Social Behavior Scales, 13th MMY
Scorable Self-Care Evaluation (Revised), 14th MMY
Social Behavior Assessment Inventory, 12th MMY
Social Competence and Behavior Evaluation, Preschool Edition, 14th MMY
Social Skills Rating System, 12th MMY
Spadafore Attention Deficit Hyperactivity Disorder Rating Scale, 14th MMY
Sutter-Eyberg Student Behavior Inventory, 11th MMY
Systematic Screening for Behavior Disorders, 13th MMY
Test of Variables of Attention, 13th MMY
Test of Variables of Attention (Version 7.03), 14th MMY
Transition Behavior Scale, 12th MMY
Transition Planning Inventory, 14th MMY
Work Readiness Profile, 14th MMY
Young Adult Behavior Checklist and Young Adult Self-Report, 14th MMY

7.14.3 Developmental Assessment

ABC Inventory to Determine Kindergarten and School Readiness (The), 10th MMY
Adolescent and Adult Psychoeducational Profile, 11th MMY
Ages and Stages Questionnaires, 14th MMY
AGS Early Screening Profiles, 12th MMY
Assessment for Persons Profoundly or Severely Impaired, 11th MMY

Assessment in Infancy: Ordinal Scales of Psychological Development, 9th MMY
Assessment in Nursery Education, 9th MMY
A. S. S. E. T. S. : A Survey of Students' Educational Talents and Skills, 10th MMY
Autism Screening Instrument for Educational Planning, Second Edition, 13th MMY
Bangs Receptive Vocabulary Checklist (The), 11th MMY
Basic School Skills Inventory, Third Edition, 9th MMY, 14th MMY
Basic School Skills Inventory—Screen, 9th MMY
Battelle Developmental Inventory, 10th MMY
Battelle Developmental Inventory Screening Test, 11th MMY
Bayley Infant Neurodevelopmental Screener, 13th MMY
Behavioral Characteristics Progression, 11th MMY
Bilingual Home Inventory, 11th MMY
Birth to Three Assessment and Intervention System, 11th MMY
Block Survey and S. L. I. D. E. (The), 12th MMY
Bracken Basic Concept Scale, 10th MMY
Bracken Basic Concept Scale—Revised, 14th MMY
BRIGANCE Early Preschool Screen for Two-Year-Old and Two-and-a-Half-Year-Old Children, 11th MMY
BRIGANCE K & 1 Screen for Kindergarten and First Grade Children [Revised], 12th MMY
Brigance Preschool Screen, 10th MMY
Bury Infant Check, 10th MMY
Caregiver's School Readiness Inventory, 12th MMY
Chicago Early Assessment and Remediation Laboratory, 11th MMY
Child Development Inventory, 13th MMY
Child Development Review, 14th MMY
Children at Risk Screener: Kindergarten and Preschool, 13th MMY
Children's Adaptive Behavior Scale, Revised, 10th MMY
Clinical Observations of Motor and Postural Skills, 14th MMY
Cognitive Control Battery, 11th MMY
Comprehensive Identification Process [Revised], 14th MMY
Dallas Pre-School Screening Test, 11th MMY
Denver II, 12th MMY
Developmental Activities Screening Inventory—II, 10th MMY
Developmental Assessment for the Severely Handicapped, 9th MMY
Developmental Assessment of Life Experiences [1986 Edition], 12th MMY
Developmental Assessment of Young Children, 14th MMY
Developmental Indicators for the Assessment of Learning, Third Edition, 14th MMY

Developmental Indicators for the Assessment of Learning—Revised/AGS Edition, 12th MMY
Developmental Observation Checklist System, 13th MMY
Developmental Profile II, 10th MMY
Developmental Tasks for Kindergarten Readiness—II, 9th MMY, 14th MMY
Diagnostic Inventory for Screening Children, 13th MMY
Early Child Development Inventory, 11th MMY
Early Intervention Developmental Profile, 13th MMY
Early Learning: Assessment and Development, 9th MMY
Early School Inventory, 13th MMY
Early Screening Inventory—Revised, 11th MMY, 14th MMY
Early Years Easy Screen, 12th MMY
ECOScales, 12th MMY
Egan Bus Puzzle Test (The), 10th MMY
Einstein Assessment of School-Related Skills, 11th MMY
Erhardt Developmental Prehension Assessment, 12th MMY
FirstSTEP: Screening Test for Evaluating Preschoolers, 13th MMY
Fisher-Landau Early Childhood Screening [Experimental Version], 14th MMY
Five P's (Parent/Professional Preschool Performance Profile) [Revised] (The), 13th MMY
Gesell Child Developmental Age Scale (The), 12th MMY
Gesell Preschool Test, 9th MMY
HELP Checklist (Hawaii Early Learning Profile), 11th MMY
Help for Special Preschoolers ASSESSMENT CHECKLIST: Ages 3-6, 11th MMY
High/Scope Child Observation Record for Ages 21/26, 14th MMY
Howell Prekindergarten Screening Test, 10th MMY
Humanics National Child Assessment Form [Revised], 11th MMY
Infant Development Inventory, 14th MMY
Infant Developmental Screening Scale, 14th MMY
Infant Screening, 10th MMY
Infant-Toddler Developmental Assessment, 13th MMY
Is This Autism? A Checklist of Behaviours and Skills for Children Showing Autistic Features, 12th MMY
Keele Pre-School Assessment Guide, 9th MMY
Kent Inventory of Developmental Skills [1996 Standardization], 14th MMY
Kindergarten Readiness Test, 12th MMY
Kindergarten Screening Inventory, 10th MMY
Learning Behaviors Scale, Research Edition, 11th MMY

Learning Inventory of Kindergarten Experiences, 12th MMY
Lexington Developmental Scales, 9th MMY
MacArthur Communicative Development Inventories, 13th MMY
Metropolitan Readiness Tests, Sixth Edition, 12th MMY, 14th MMY
Milani-Comparetti Motor Development Screening Test, 11th MMY
Miller Assessment for Preschoolers, 9th MMY
Missouri Kindergarten Inventory of Developmental Skills, Alternate Form, 11th MMY
Movement Assessment Battery for Children, 14th MMY
Mullen Scales of Early Learning: AGS Edition, 14th MMY
Neonatal Behavioral Assessment Scale, 3rd Edition, 14th MMY
Neurobehavioral Assessment of the Preterm Infant, 13th MMY
Oral-Motor/Feeding Rating Scale, 12th MMY
PACE, 12th MMY
Parents' Evaluation of Developmental Status, 14th MMY
Pediatric Early Elementary Examination—II, 13th MMY
Pediatric Evaluation of Disability Inventory, 14th MMY
Pediatric Examination of Educational Readiness at Middle Childhood [Revised], 10th MMY,14th MMY
Pediatric Extended Examination at Three, 11th MMY
Phelps Kindergarten Readiness Scale, 12th MMY
Pictorial Scale of Perceived Competence and Social Acceptance for Young Children (The),11th MMY
Play Observation Scale (The), 13th MMY
Pragmatics Profile of Early Communication Skills (The), 12th MMY
Pre-School Behavior Checklist, 11th MMY
Preschool Development Inventory, 13th MMY
Preschool Developmental Profile, 13th MMY
Preschool Evaluation Scale (The), 13th MMY
Preschool Language Scale, 13th MMY
Preschool Screening Test, 12th MMY
Preschool Skills Test, 14th MMY
Pre-Verbal Communication Schedule, 11th MMY
Printing Performance School Readiness Test, 12th MMY
Program for the Acquisition of Language with the Severely Impaired, 11th MMY
Psychoeducational Profile Revised, 12th MMY
Revised BRIGANCE Diagnostic Inventory of Early Development, 12th MMY
Revised Denver Prescreening Developmental Questionnaire, 12th MMY
Schedule of Growing Skills (The), 11th MMY

Screening Test for Educational Prerequisite Skills, 12th MMY
Sequenced Inventory of Communication Development, Revised Edition, 10th MMY
Southern California Ordinal Scales of Development (The), 10th MMY
Spatial Orientation Memory Test, 11th MMY
Speech-Ease Screening Inventory (K-1), 12th MMY
Study of Children's Learning Behaviors, Research Edition, 11th MMY
Survey of Students' Educational Talents and Skills. *See* A. S. S. E. T. S.: A Survey of Students' Educational Talents and Skills
Symbolic Play Test, Second Edition, 12th MMY
System to Plan Early Childhood Services, 12th MMY
TARC Assessment System (The), 11th MMY
Teacher's School Readiness Inventory, 12th MMY
Test of Kindergarten/First Grade Readiness Skills, 13th MMY
Test of Sensory Functions in Infants, 11th MMY
Toddler and Infant Motor Evaluation, 14th MMY
Transdisciplinary Play-Based Assessment, Revised Edition, 13th MMY
Vineland Social-Emotional Early Childhood Scales, 14th MMY
Vulpe Assessment Battery—Revised, 14th MMY
Wisconsin Behavior Rating Scale, 11th MMY

7.14.4 Education

Academic Advising Inventory, 10th MMY
Achievement Identification Measure—Teacher Observation, 11th MMY
ACT Evaluation/Survey Service (The), 12th MMY
ACT Study Skills Assessment and Inventory (College Edition), 12th MMY
Adaptive Style Inventory, 13th MMY
Analytic Learning Disability Assessment, 9th MMY
Assessment of School Needs for Low-Achieving Students: Staff Survey, 11th MMY
Class Activities Questionnaire (The), 9th MMY
Classroom Communication Screening Procedure for Early Adolescents: A Handbook for Assessment and Intervention, 11th MMY
Classroom Environment Scale, Second Edition, 10th MMY
College Major Interest Inventory, 12th MMY
College Outcome Measures Program, 13th MMY
College Student Experiences Questionnaire, 10th MMY
Community College Goals Inventory, 9th MMY
Community College Student Experiences Questionnaire, 12th MMY
Community College Student Experiences Questionnaire, Second Edition, 14th MMY

Comprehensive Assessment of School Environments, 11th MMY
Comprehensive Assessment of School Environments Information Management System, 14th MMY
Computer Literacy and Computer Science Tests, 11th MMY
Content Mastery Examinations for Educators, 12th MMY
Cooperative Institutional Research Program (The), 9th MMY
Course Finder 2000, 13th MMY
Course-Faculty Instrument, 9th MMY
Creative Styles Inventory, 13th MMY
DABERON-2: Screening for School Readiness, 11th MMY
DeLong Interest Inventory, 11th MMY
Dimensions of Excellence Scales, 11th MMY
Dimensions of Excellence Scales[1991 Edition], 14th MMY
Dropout Prediction & Prevention, 12th MMY
Early Childhood Environment Rating Scale—Revised, 9th MMY
Early Childhood Environment Rating Scale—Revised Edition, 14th MMY
Educational Administrator Effectiveness Profile, 10th MMY
Educational Interest Inventory, 11th MMY
Educational Leadership Practices Inventory, 11th MMY
Educational Process Questionnaire, 11th MMY
Effective School Battery (The), 10th MMY
Evaluating Educational Programs for Intellectually Gifted Students, 10th MMY
Gifted Program Evaluation Survey (The), 12th MMY
How a Child Learns, 9th MMY
Hudson Education Skills Inventory, 11th MMY
ICES: Instructor and Course Evaluation System, 9th MMY
IDEA, 9th MMY
IDEA Student Ratings of Instruction, 14th MMY
Independent School Entrance Examination, 11th MMY
Individualised Classroom Environment Questionnaire, 11th MMY
Instructional Environment System—II (The), 13th MMY
Instructional Leadership Evaluation and Development Program (ILEAD), 12th MMY
Instructor and Course Evaluation System. *See* ICES: Instructor and Course Evaluation System
Inventory of Learning Processes, 13th MMY
Learning and Study Strategies Inventory, 11th MMY
Learning and Study Strategies Inventory—High School Version, 13th MMY
Learning Preference Scales, 13th MMY

Learning Style Inventory [Price Systems, Inc.], 13th MMY
Learning Styles and Strategies, 11th MMY
Learning Styles Inventory [Educational Activities, Inc.], 14th MMY
Learning Styles Inventory [Piney Mountain Press, Inc.], 12th MMY
Light's Retention Scale [Revised Edition, 1991], 11th MMY
Measure of Questioning Skills, 14th MMY
Middle Grades Assessment Program, 14th MMY
Minnesota School Attitude Survey, 9th MMY
Motivated Strategies for Learning Questionnaire, 13th MMY
Noncognitive Variables and Questionnaire, 14th MMY
NTE Programs, 9th MMY
Oetting Michaels Anchored Ratings for Therapists, 14th MMY
Our Class and Its Work, 10th MMY
PEEK—Perceptions, Expectations, Emotions, and Knowledge About College, 14th MMY
Performance Levels of a School Program Survey, 10th MMY
Prescriptive Teaching Series, 11th MMY
Program Self-Assessment Service, 10th MMY
Rating Inventory for Screening Kindergartners (A), 12th MMY
Scales for Effective Teaching (The), 13th MMY
Scales for Predicting Successful Inclusion, 14th MMY
School Archival Records Search, 13th MMY
School Assessment Survey, 12th MMY
School Effectiveness Questionnaire, 13th MMY
School Environment Preference Survey, 10th MMY
School Situation Survey, 12th MMY
School Social Skills Rating Scale, 12th MMY
School-Age Care Environment Rating Scale, 14th MMY
SCREEN [Senf-Comrey Ratings of Extra Educational Need], 11th MMY
Screening Children for Related Early Educational Needs, 11th MMY
Secondary School Admission Test, 11th MMY
Student Goals Exploration, 13th MMY
Student Rights Scales, 10th MMY
Study Attitudes and Methods Survey [Revised Short Form], 14th MMY
Styles of Training Index, 10th MMY
Survey of Student Opinion of Instruction, 10th MMY
Surveys of Problem-Solving and Educational Skills, 11th MMY
Teacher Evaluation Rating Scales, 11th MMY

Teacher Evaluation Scale, 12th MMY
Teacher Role Survey (Revised), 10th MMY
Test of Attitude Toward School, 11th MMY
Thinking About My School, 10th MMY
Vocational Learning Styles, 12th MMY

7.14.5 English

ACO. *See* Assessment of Conceptual Organization (ACO): Improving, Writing, Thinking, and Reading Skills
Adolescent Language Screening Test, 10th MMY
Adult Language Assessment Scales, 14th MMY
Adult Rating of Oral English, 14th MMY
Alberta Essay Scales: Models, 9th MMY
Alphabet Mastery, 9th MMY
Analyzing the Communication Environment, 14th MMY
Assessing Motivation to Communicate, 14th MMY
Assessment of Children's Language Comprehension, 1983 Revision, 10th MMY
Assessment of Conceptual Organization (ACO): Improving, Writing, Thinking, and Reading Skills, 12th MMY
Bankson Language Test—2, 11th MMY
Basic English Skills Test, 11th MMY
Basic Inventory of Natural Language, 9th MMY
Basic Language Concepts Test, 10th MMY
Beery Picture Vocabulary Test and Beery Picture Vocabulary Screening Series, 12th MMY
Bristol Language Development Scales, 12th MMY
British Ability Scales: Spelling Scale, 12th MMY
California Achievement Tests Writing Assessment System, 11th MMY
CAP Assessment of Writing, 12th MMY
CAT/5 Listening and Speaking Checklist, 13th MMY
Clark-Madison Test of Oral Language, 10th MMY
Classroom Communication Skills Inventory: Listening and Speaking Checklist (The), 13th MMY
Clinical Evaluation of Language Fundamentals, Third Edition—Observational Rating Scales, 14th MMY
Clinical Evaluation of Language Fundamentals-3 Screening Test, 13th MMY
Communication Abilities Diagnostic Test, 12th MMY
Communication Skills Profile, 14th MMY

Competent Speaker Speech Evaluation Form (The), 14th MMY
Comprehension of Oral Language, 9th MMY
Comprehensive Receptive and Expressive Vocabulary Test, 13th MMY
Comprehensive Receptive and Expressive Vocabulary Test—Adult, 14th MMY
Computer-Based Test of English as a Foreign Language and Test of Written English, 14th MMY
Conversational Skills Rating Scale: An Instructional Assessment of Interpersonal Competence(The), 14th MMY
CTB Writing Assessment System, 13th MMY
Descriptive Tests of Language Skills, 11th MMY
Diagnostic Achievement Test in Spelling (A), 10th MMY
Diagnostic Screening Test: Language, Second Edition, 9th MMY
Diagnostic Screening Test: Spelling, Third Edition, 9th MMY
Diagnostic Spelling Potential Test, 9th MMY
Diagnostic Spelling Test, 9th MMY
Dos Amigos Verbal Language Scales, 14th MMY
Early Language Milestone Scale, Second Edition, 13th MMY
English as a Second Language Oral Assessment, Revised, 11th MMY
English Language Skills Assessment in a Reading Context, 11th MMY
English Skills Assessment, 9th MMY
ERB Writing Assessment, 13th MMY
ESL/Adult Literacy Scale, 11th MMY
ETS Tests of Applied Literacy Skills, 13th MMY
Expressive Vocabulary Test, 14th MMY
Figurative Language Interpretation Test, 12th MMY
First Words and First Sentences Tests, 14th MMY
Fullerton Language Test for Adolescents, Second Edition (The), 10th MMY
Graded Word Spelling Test, 12th MMY
Graded Word Spelling Test, Second Edition, 14th MMY
Grammatical Analysis of Elicited Language—Simple Sentence Level, Second Edition, 11th MMY
Group Literacy Assessment, 9th MMY
Henderson-Moriarty ESL/Literacy Placement Test, 10th MMY
Houston Test for Language Development, Revised Edition (The), 11th MMY
IDEA Oral Language Proficiency Test, 14th MMY
IDEA Reading and Writing Proficiency Test, 13th MMY
Independent Mastery Testing System for Writing Skills, 11th MMY
Informal Writing Inventory, 11th MMY

Integrated Literature and Language Arts Portfolio Program, 12th MMY
Integrated Writing Test, 14th MMY
Inventory of Language Abilities, 9th MMY
Inventory of Language Abilities, Level II, 10th MMY
Kindergarten Language Screening Test, Second Edition, 14th MMY
Language Arts Assessment Portfolio, 13th MMY
Language Assessment Battery, 9th MMY
Language Assessment Scales—Oral, 12th MMY
Language Assessment Scales, Reading and Writing, 12th MMY
Language Facility Test, 9th MMY
Language Imitation Test, 9th MMY
Language Processing Test, 10th MMY
Language Proficiency Test, 9th MMY
Language Sampling, Analysis, and Training, Revised Edition, 9th MMY
Language Sampling, Analysis, and Training—Third Edition, 14th MMY
Learning Through Listening, 9th MMY
Listen Up: Skills Assessment, 14th MMY
Listening Comprehension Test, 12th MMY
Listening Styles Profile, 14th MMY
MAC Checklist for Evaluating, Preparing, and/or Improving Standardized Tests for Limited English Speaking Students, 10th MMY
Mather-Woodcock Group Writing Tests, 14th MMY
Michigan English Language Assessment Battery, 14th MMY
Michigan Prescriptive Program in English, 9th MMY
Miller-Yoder Language Comprehension Test (Clinical Edition), 11th MMY
Multilevel Informal Language Inventory, 10th MMY
Oral and Written Language Scales: Listening Comprehension and Oral Expression, 14th MMY
Oral and Written Language Scales: Written Expression, 14th MMY
Oral Language Evaluation, Second Edition, 12th MMY
Parallel Spelling Tests, 12th MMY
Parallel Spelling Tests, Second Edition, 14th MMY
PRE-LAS English, 11th MMY
Preliminary Test of English as a Foreign Language, 11th MMY
Primary Language Screen (The), 12th MMY
Progressive Achievement Tests of Listening Comprehension[Revised], 13th MMY
Quick Informal Assessment, 13th MMY
Quick Phonics Survey, 12th MMY

Quick Spelling Inventory, 11th MMY
Receptive Oral Language Inventory, 14th MMY
Renfrew Bus Story (The), 14th MMY
Reynell Developmental Language Scales [U. S. Edition], 12th MMY
Reynell Developmental Language Scales III: The University of Reading Edition [British Edition], 14th MMY
Rhode Island Test of Language Structure, 10th MMY
Second Language Oral Test of English, 11th MMY
Self-Assessment in Writing Skills, 12th MMY
Sentence Comprehension Test, Revised Edition, 11th MMY
Slingerland College-Level Screening for the Identification of Language Learning Strengths and Weaknesses, 12th MMY
Spanish Structured Photographic Expressive Language Test, 11th MMY
Spellmaster, 11th MMY
Stanford Writing Assessment Program, Second Edition, 13th MMY
Stanford Writing Assessment Program, Third Edition, 14th MMY
Teacher Assessment of Grammatical Structures, 10th MMY
Test for Auditory Comprehension of Language—Third Edition, 14th MMY
Test of Adolescent and Adult Language, Third Edition, 13th MMY
Test of Adolescent/Adult Word Finding, 12th MMY
Test of Children's Language: Assessing Aspects of Spoken Language, Reading, and Writing, 14th MMY
Test of Early Language Development, Second Edition, 12th MMY
Test of Early Language Development, Third Edition, 14th MMY
Test of Early Written Language, Second Edition, 13th MMY
Test of English for International Communication (TOEIC), 11th MMY
Test of English Proficiency Level, 11th MMY
Test of Language Competence—Expanded Edition, 11th MMY
Test of Language Development—Intermediate, Third Edition, 11th MMY, 14th MMY
Test of Language Development—Primary, Third Edition, 11th MMY, 14th MMY
Test of Word Finding, 11th MMY
Test of Word Finding in Discourse, 12th MMY
Test of Word Knowledge, 13th MMY
Test of Written Expression, 13th MMY
Test of Written Language—Third Edition, 13th MMY
Test of Written Spelling, Third Edition, 13th MMY
TOEIC. *See* Test of English for International Communication (TOEIC)
Utah Test of Language Development—3, 11th MMY

Woodcock Language Proficiency Battery—Revised, 12th MMY
Woodcock-Munoz Language Survey, 13th MMY
Writing Process Test, 13th MMY
Written Expression Test (The), 10th MMY
Written Language Assessment, 11th MMY

7.14.6 Fine Arts

Advanced Measures of Music Audiation, 12th MMY
Group Tests of Musical Abilities, 12th MMY
Instrument Timbre Preference Test, 10th MMY
Iowa Tests of Music Literacy, Revised, 13th MMY
Measures of Musical Abilities, 10th MMY
Music Achievement Tests 1, 2, 3, and 4, 12th MMY
Musical Aptitude Profile [1988 Revision], 12th MMY

7.14.7 Foreign Languages

Ber-Sil Spanish Test (The), 10th MMY
Bilingual Syntax Measure, 9th MMY
Chinese Proficiency Test, 13th MMY
Chinese Speaking Test, 13th MMY
El Circo, 9th MMY
Hausa Speaking Test, 14th MMY
Hebrew Speaking Test, 13th MMY
Indonesian Speaking Test, 13th MMY
LOTE Reading and Listening Tests, 12th MMY
Portuguese Speaking Test, 12th MMY
Spanish Computerized Adaptive Placement Exam (A), 12th MMY

7.14.8 Intelligence and Scholastic Aptitude

ACCUPLACER: Computerized Placement Tests, 13th MMY
ACER Advanced Test B40, 9th MMY
ACER Advanced Test B90: New Zealand Edition, 12th MMY
ACER Advanced Tests AL and AQ (Second Edition) and BL-BQ, 9th MMY
ACER Intermediate Test G, 9th MMY
ACER Test of Reasoning Ability, 12th MMY
ACER Word Knowledge Test, 11th MMY
Adaptation of the Wechsler Preschool and Primary Scale of Intelligence for Deaf Children (An), 11th MMY

AH1 Forms X and Y, 9th MMY
AH6 Group Tests of High Level Intelligence, 10th MMY
Alternate Uses, 9th MMY
Arlin Test of Formal Reasoning, 9th MMY
Assessment of Individual Learning Style: The Perceptual Memory Task, 11th MMY
Bateria Woodcock-Munoz—Revisada, 13th MMY
Bilingual Verbal Ability Tests, 14th MMY
Boehm Test of Basic Concepts—Preschool Version, 11th MMY
Boehm Test of Basic Concepts—Revised, 10th MMY
British Ability Scales (The), 9th MMY
California Critical Thinking Dispositions Inventory (The), 12th MMY
California Critical Thinking Skills Test, 12th MMY
Canadian Cognitive Abilities Test, Form 7, 12th MMY
Canadian Test of Cognitive Skills, 13th MMY
Children's Abilities Scales, 10th MMY
Children's Category Test, 13th MMY
CID Preschool Performance Scale, 12th MMY
Cognitive Abilities Scale, 10th MMY
Cognitive Abilities Test, Form 5, 13th MMY
Cognitive Diagnostic Battery, 9th MMY
Cognitive Observation Guide, 12th MMY
Comprehensive Adult Student Assessment System, 13th MMY
Comprehensive Scales of Student Abilities: Quantifying Academic Skills and School-Related Behavior Through the Use of Teacher Judgments, 13th MMY
Comprehensive Test of Nonverbal Intelligence, 13th MMY
Continuous Visual Memory Test [Revised], 12th MMY, 14th MMY
Cornell Critical Thinking Tests (The), 11th MMY
Cree Questionnaire, 9th MMY
Critical Reasoning Tests, 12th MMY
Das-Naglieri Cognitive Assessment System, 14th MMY
Deductive Reasoning Test, 9th MMY
Detroit Tests of Learning Aptitude, Fourth Edition, 12th MMY, 14th MMY
Detroit Tests of Learning Aptitude—Adult, 12th MMY
Detroit Tests of Learning Aptitude—Primary, Second Edition, 12th MMY
Developing Cognitive Abilities Test [Second Edition], 11th MMY
Drumcondra Verbal Reasoning Test 1, 11th MMY
Edinburgh Picture Test, 12th MMY
Ennis-Weir Critical Thinking Essay Test, 10th MMY

Expressive One-Word Picture Vocabulary Test, Revised, 12th MMY
Extended Merrill-Palmer Scale (The), 12th MMY
Fuld Object-Memory Evaluation, 9th MMY
General Ability Measure for Adults, 14th MMY
General Processing Inventory, 14th MMY
Gifted and Talented Scale, 10th MMY
Gifted Evaluation Scale, 12th MMY
Gifted Evaluation Scale, Second Edition, 14th MMY
Graduate and Managerial Assessment, 10th MMY
Graduate Management Admission Test, 9th MMY
Graduate Record Examinations—General Test (The), 9th MMY
Hammill Multiability Intelligence Test, 14th MMY
Harding Skyscraper, 9th MMY
Henmon-Nelson Ability Test, Canadian Edition, 12th MMY
Jenkins Non-Verbal Test, 1986 Revision, 12th MMY
Kaufman Adolescent and Adult Intelligence Test, 12th MMY
Kaufman Assessment Battery for Children, 9th MMY
Kaufman Brief Intelligence Test, 12th MMY
Kaufman Infant and Preschool Scale, 9th MMY
Khatena-Morse Multitalent Perception Inventory, 13th MMY
Knox's Cube Test, 9th MMY
Learning Ability Profile, 9th MMY
Learning Disability Evaluation Scale (Renormed), 12th MMY, 14th MMY
Learning Disability Rating Procedure, 9th MMY
Learning Efficiency Test-II (1992 Revision), 12th MMY
Leiter International Performance Scale—Revised, 14th MMY
Matrix Analogies Test, 10th MMY
McCarthy Scales of Children's Abilities, 9th MMY
McCarthy Screening Test, 9th MMY
MD5 Mental Ability Test, 10th MMY
Memory Assessment Scales, 12th MMY
Merrill-Palmer Scale of Mental Tests, 9th MMY
MIDAS: Multiple Intelligence Developmental Assessment Scales (The), 14th MMY
Mill Hill Vocabulary Scale, 13th MMY
Miller Analogies Test, 12th MMY
Modified Vygotsky Concept Formation Test, 12th MMY
Multiple Intelligence Developmental Asscssment Scales. *See* MIDAS: Multiple Intelligence Developmental Assessment Scales (The)

Naglieri Nonverbal Ability Test, 14th MMY
New Jersey Test of Reasoning Skills—Form B, 11th MMY
Non-Reading Intelligence Tests, Levels 1-3, 12th MMY
Non-Verbal Ability Tests, 12th MMY
Otis-Lennon School Ability Test, Seventh Edition, 11th MMY, 14th MMY
Paper and Pencil Games, 14th MMY
Peabody Picture Vocabulary Test-Ⅲ, 9th MMY, 14th MMY
Preschool and Kindergarten Interest Descriptor, 10th MMY
Primary Test of Cognitive Skills, 12th MMY
Process Skills Rating Scales, 12th MMY
Randt Memory Test, 10th MMY
Receptive One-Word Picture Vocabulary Test, 10th MMY
Receptive One-Word Picture Vocabulary Test: Upper Extension, 11th MMY
Recognition Memory Test, 10th MMY
Riley Inventory of Basic Learning Skills, 12th MMY
Schaie-Thurstone Adult Mental Abilities Test, 10th MMY
Scholastic Abilities Test for Adults, 12th MMY
Screening Assessment for Gifted Elementary Students, 10th MMY
Screening Assessment for Gifted Elementary Students—Primary, 12th MMY
Selby MillSmith Adaptive Ability Tests, 13th MMY
Shipley Institute of Living Scale, 11th MMY
Silver Drawing Test of Cognition and Emotion[Third Edition Revised], 14th MMY
Slosson Full-Range Intelligence Test, 14th MMY
Slosson Intelligence Test [1991 Edition], 12th MMY
Stanford-Binet Intelligence Scale, Fourth Edition, 10th MMY
Structure of Intellect Learning Abilities Test, 10th MMY
Swanson-Cognitive Processing Test, 13th MMY
Teacher Observation Scales for Identifying Children with Special Abilities, 14th MMY
Test of Cognitive Skills, Second Edition, 13th MMY
Test of Learning Ability, 12th MMY
Test of Memory and Learning, 13th MMY
Test of Nonverbal Intelligence, Third Edition, 11th MMY, 14th MMY
Test of Relational Concepts, 11th MMY
Test on Appraising Observations Tests of Adult Basic Education Work-Related Problem Solving, 13th MMY
Thinking Creatively in Action and Movement, 9th MMY
Thinking Creatively With Sounds and Words, Research Edition, 9th MMY
Torrance Tests of Creative Thinking, 9th MMY

Universal Nonverbal Intelligence Test, 14th MMY
USES General Aptitude Test Battery for the Deaf, 13th MMY
Watson-Glaser Critical Thinking Appraisal, Form S, 13th MMY
Wechsler Abbreviated Scale of Intelligence, 14th MMY
Wechsler Adult Intelligence Scale—Third Edition, 14th MMY
Wechsler Intelligence Scale for Children, Third Edition, 12th MMY
Wechsler Memory Scale III, 11th MMY, 14th MMY
Wechsler Preschool and Primary Scale of Intelligence—Revised, 11th MMY
Wide Range Assessment of Memory and Learning, 11th MMY
Wiesen Test of Mechanical Aptitude (The), 14th MMY
Wisconsin Card Sorting Test, Revised and Expanded, 14th MMY
Wonderlic Personnel Test and Scholastic Level Exam, 11th MMY, 14th MMY
Woodcock-Johnson Psycho-Educational Battery—Revised, 12th MMY

7.14.9 Mathematics

American High School Mathematics Examination, 11th MMY
American Invitational Mathematics Examination, 11th MMY
American Junior High School Mathematics Examination, 11th MMY
Arithmetic Skills Assessment Test, 14th MMY
Assessment in Mathematics, 9th MMY
Basic Number Diagnostic Test, 9th MMY
Basic Number Screening Test, 9th MMY
Booker Profiles in Mathematics: Numeration and Computation, 13th MMY
Buswell-John Diagnostic Test for Fundamental Processes in Arithmetic, 9th MMY
California Diagnostic Mathematics Tests, 11th MMY
Class Achievement Test in Mathematics, 11th MMY
Collis-Romberg Mathematical Problem Solving Profiles, 12th MMY
Descriptive Tests of Mathematics Skills, 11th MMY
Diagnostic Mathematics Profiles, 12th MMY
Diagnostic Screening Test: Math, Third Edition, 9th MMY
Diagnostic Test of Arithmetic Strategies, 9th MMY
Early Mathematics Diagnostic Kit, 11th MMY
ENRIGHT Diagnostic Inventory of Basic Arithmetic Skills, 10th MMY
Group Mathematics Test, Second Edition, 9th MMY
Independent Mastery Testing System for Math Skills, 10th MMY
Iowa Algebra Aptitude Test, Fourth Edition, 12th MMY
KeyMath Revised: A Diagnostic Inventory of Essential Mathematics, 11th MMY
KeyMath Revised: A Diagnostic Inventory of Essential Mathematics [1998 Normative

Update],14th MMY
Mathematical Olympiads, 11th MMY
Mathematics Anxiety Rating Scale, 11th MMY
Mathematics Attitude Inventory, 9th MMY
Mathematics Competency Test, 14th MMY
Mathematics 7, 11th MMY
Mathematics Topic Pre-Tests, 13th MMY
Practical Maths Assessments, 12th MMY
Profiles of Problem Solving, 13th MMY
Progressive Achievement Test of Mathematics [Revised], 12th MMY
Progressive Achievement Tests in Mathematics, 11th MMY
Ramsay Corporation Job Skills—Office Arithmetic Test, 13th MMY
Sequential Assessment of Mathematics Inventories: Standardized Inventory, 11th MMY
Stanford Diagnostic Mathematics Test, Fourth Edition, 13th MMY
Target Mathematics Tests, 14th MMY
Test of Cognitive Style in Mathematics, 12th MMY
Test of Early Mathematics Ability, Second Edition, 11th MMY
Test of Mathematical Abilities, Second Edition, 13th MMY
Test of Mathematical Abilities for Gifted Students, 14th MMY

7.14.10 Miscellaneous

ACCESS—A Comprehensive Custody Evaluation Standard System, 14th MMY
Achievement Identification Measure, 10th MMY
Ackerman-Schoendorf Scales for Parent Evaluation of Custody, 12th MMY
ACT Study Power Assessment and Inventory, 12th MMY
Adaptive Behavior: Street Survival Skills Questionnaire, 11th MMY
Adaptive Functioning Index, 9th MMY
Adolescent Diagnostic Interview, 12th MMY
Adult Growth Examination, 11th MMY
Aggregate Neurobehavioral Student Health and Educational Review. *See* ANSER System-Aggregate Neurobehavioral Student Health and Educational Review
Ahr's Individual Development Survey, 9th MMY
Alcohol Clinical Index, 13th MMY
Alcohol Use Disorders Identification Test, 14th MMY
American Drug and Alcohol Survey (The), 12th MMY
Ann Arbor Learning Inventory, 11th MMY
ANSER System-Aggregate Neurobehavioral Student Health and Educational Review (The),9th MMY

Arizona Battery for Communication Disorders of Dementia (The), 12th MMY
ASIST: A Structured Addictions Assessment Interview for Selecting Treatment, 13th MMY
Assessment of Chemical Health Inventory, 13th MMY
Assessment of Interpersonal Relations, 13th MMY
Assessment of Living Skills and Resources, 13th MMY
Assessment of Parenting Skills: Infant and Preschooler, 14th MMY
Auditory Selective Attention Test, 13th MMY
Basic Tests Series, 11th MMY
Behavioral Assessment of Pain Questionnaire, 12th MMY
Bloomer Learning Test, 9th MMY
Braille Assessment Inventory, 14th MMY
Bricklin Perceptual Scales, 11th MMY
Career Assessment Inventories for the Learning Disabled, 9th MMY
Checklist for Child Abuse Evaluation, 12th MMY
Checklist/Guide to Selecting a Small Computer, 9th MMY
Child Care Inventory, 12th MMY
Childhood Trauma Questionnaire, 14th MMY
Children's Version of The Family Environment Scale, 10th MMY
Clinical Rating Scale, 12th MMY
Clinical Support System Battery, 12th MMY
Cognitive Skills Assessment Battery, Second Edition, 9th MMY
Comprehensive Custody Evaluation Standard System. *See* ACCESS—A Comprehensive Custody Evaluation Standard System
Correctional Institutions Environment Scale, Second Edition, 14th MMY
Courtship Analysis (A), 9th MMY
Creativity Assessment Packet, 9th MMY
DANTES Subject Standardized Tests, 11th MMY
Determining Needs in Your Youth Ministry, 12th MMY
Diagnosing Organizational Culture, 13th MMY
Diagnosis and Remediation of Handwriting Problems, 10th MMY
Diagnostic Test of Library Skills—Advanced Edition, 13th MMY
Driver Risk Inventory, 12th MMY
Driver Risk Inventory—II, 14th MMY
Drug Use Questionnaire, 13th MMY
Drug-Taking Confidence Questionnaire, 14th MMY
Dyslexia Determination Test, 9th MMY
Early Childhood Physical Environment Observation Schedules and Rating Scales, 13th

MMY
Eating Inventory, 13th MMY
Educational Development Series (1992 Edition), 9th MMY
Endeavor Instructional Rating System, 9th MMY
Enhanced ACT Assessment, 12th MMY
Entrepreneurial Style and Success Indicator, 13th MMY
Facial Action Coding System, 12th MMY
Family Assessment Form: A Practice-Based Approach to Assessing Family Functioning, 14th MMY
Family Assessment Measure Version III, 14th MMY
Family Day Care Rating Scale, 11th MMY
Family Environment Scale [Third Edition Manual], 14th MMY
Family Environment Scale, Second Edition, 12th MMY
Family Risk Scales, 13th MMY
Fast Health Knowledge Test, 1986 Revision, 10th MMY
Food Choice Inventory, 10th MMY
Functional Fitness Assessment for Adults over 60 Years, 13th MMY
Functional Performance Record, 12th MMY
Functional Time Estimation Questionnaire, 12th MMY
Golombok Rust Inventory of Marital State (The), 12th MMY
Goodman Lock Box (The), 9th MMY
Grandparent Strengths and Needs Inventory, 13th MMY
Gregorc Style Delineator, 12th MMY
Group Environment Scale, Second Edition, 10th MMY
GROW—The Marriage Enrichment Program, 10th MMY
Health and Daily Living, 10th MMY
Health Problems Checklist, 10th MMY
Herrmann Brain Dominance Instrument, 11th MMY
Herrmann Brain Dominance Instrument [Revised], 14th MMY
Home Environment Questionnaire, HEQ-2R and HEQ-1R, 10th MMY
Home Observation for Measurement of the Environment, 9th MMY
Home Screening Questionnaire, 9th MMY
How Am I Doing? A Self-Assessment for Child Caregivers, 13th MMY
Identi-Form System for Gifted Programs (The), 10th MMY
Individual Service Strategy Portfolio, 13th MMY
Infant/Toddler Environment Rating Scale, 12th MMY
Influence Strategies Exercise, 13th MMY
Interpersonal Conflict Scale, 9th MMY

Inventory for Client and Agency Planning, 10th MMY
Inventory of Drinking Situations, 13th MMY
Inventory of Drug-Taking Situations, 14th MMY
I-SPEAK Your Language: A Survey of Personal Styles, 13th MMY
Jordan Left-Right Reversal Test (1990 Edition), 12th MMY
Kilmann-Saxton Culture-Gap Survey [Revised], 13th MMY
Kohlman Evaluation of Living Skills (The), 14th MMY
Krantz Health Opinion Survey, 9th MMY
Language-Structured Auditory Retention Span Test, 11th MMY
Lawrence Psychological Forensic Examination (The), 10th MMY
Learning Environment Inventory, 9th MMY
Level of Service Inventory—Revised (The), 14th MMY
Life Experiences Checklist, 12th MMY
Life Stressors and Social Resources Inventory—Adult Form, 13th MMY
Life Stressors and Social Resources Inventory—Youth Form, 13th MMY
Lifestyle Assessment Questionnaire, 13th MMY
Listening Comprehension, 9th MMY
Love Attitudes Inventory (A), 9th MMY
Marital Check-Up Kit, 13th MMY
Marital Evaluation Checklist, 10th MMY
Marriage and Family Attitude Survey (The), 11th MMY
Marriage Evaluation (A), 9th MMY
Marriage Expectation Inventories (The), 9th MMY
Marriage Role Expectation Inventory (A), 9th MMY
Matching Person and Technology, 14th MMY
Mathematics Self-Efficacy Scale, 14th MMY
Medical Ethics Inventory, 10th MMY
Michigan Alcoholism Screening Test, 14th MMY
Military Environment Inventory, 11th MMY
Multidimensional Health Profile, 14th MMY
Multiphasic Environmental Assessment Procedure, 12th MMY
My Class Inventory, 9th MMY
Older Persons Counseling Needs Survey, 14th MMY
Opinions About Deaf People Scale, 14th MMY
Organizational Climate Exercise II, 13th MMY
Organizational Culture Inventory, 12th MMY
Pain Assessment Battery, Research Edition, 14th MMY
Parent Perception of Child Profile, 12th MMY

Parent-Adolescent Communication Scale, 12th MMY
Parent/Family Involvement Index, 10th MMY
Parenting Satisfaction Scale, 14th MMY
Perceptions of Parental Role Scales, 11th MMY
Personal Assessment of Intimacy in Relationships, 11th MMY
Personal Communication Plan (The), 12th MMY
Personal Experience Inventory for Adults, 13th MMY
Personal Experience Screening Questionnaire, 12th MMY
Personal Relationship Inventory, 12th MMY
Pollack-Brandcn Inventory, 11th MMY
PREPARE/ENRICH, 14th MMY
Prevocational Assessment and Curriculum Guide (The), 12th MMY
Prison Inmate Inventory, 14th MMY
Productivity Environmental Preference Survey, 13th MMY
Project Implementation Profile, 13th MMY
Psychosocial Pain Inventory [Revised], 14th MMY
Quality of Life Inventory, 14th MMY
Quality Potential Assessment, 14th MMY
Quickview Social History, 13th MMY
RAND-36 Health Status Inventory, 14th MMY
Rating Scale of Communication in Cognitive Decline, 12th MMY
Recovery Attitude and Treatment Evaluator, 14th MMY
Rehabilitation Compliance Scale, 14th MMY
Relating to Each Other: A Questionnaire for Students, 13th MMY
Relating with Colleagues: A Questionnaire for Faculty Members, 13th MMY
Retirement Activities Card Sort Planning Kit, 13th MMY
Rogers Criminal Responsibility Assessment Scales, 10th MMY
SAQ-Adult Probation [Substance Abuse Questionnaire], 12th MMY
Scale of Marriage Problems: Revised, 14th MMY
Sex-Role Egalitarianism Scale, 13th MMY
SF-36 Health Survey, 14th MMY
Shoplifting Inventory, 14th MMY
Signals Listening Test, 10th MMY
Situational Confidence Questionnaire, 13th MMY
Small Business Assessment, 13th MMY
Smell Identification Test [Revised], 14th MMY
Social and Prevocational Information Battery—Revised, 11th MMY
Softball Skills Test, 12th MMY

Standardized Test of Computer Literacy and Computer Anxiety Index (Version AZ), Revised, 10th MMY
START—Strategic Assessment of Readiness for Training, 14th MMY
Stress Indicator & Health Planner, 13th MMY
Structured Addictions Assessment Interview for Selecting Treatment. *See* ASIST: A Structured Addictions Assessment Interview for Selecting Treatment
Study Habits Evaluation and Instruction Kit, 11th MMY
Styles of Conflict Inventory, 13th MMY
Substance Abuse Questionnaire. *See* SAQ-Adult Probation [Substance Abuse Questionnaire]
Substance Abuse Relapse Assessment, 13th MMY
Substance Abuse Screening Test, 13th MMY
Substance Use Disorder Diagnosis Schedule, 13th MMY
Survey of Functional Adaptive Behaviors, 11th MMY
Survey of Organizational Stress, 11th MMY
Survey of Organizations, 13th MMY
Swassing-Barbe Modality Index, 9th MMY
System of Multicultural Pluralistic Assessment, 9th MMY
Team Effectiveness and Meeting Strategy, 10th MMY
Tennis Skills Test Manual, 12th MMY
Test of Auditory Analysis Skills, 11th MMY
Test of Legible Handwriting, 11th MMY
TestWell: Health Risk Appraisal, 13th MMY
TestWell: Wellness Inventory, 13th MMY
TestWell: Wellness Inventory—College Version, 13th MMY
Time Use Analyzer, 14th MMY
Time-Sample Behavioral Checklist (The), 12th MMY
Uniform Child Custody Evaluation System, 13th MMY
University Residence Environment Scale [Revised], 14th MMY
Vineland Adaptive Behavior Scales, 10th MMY
Ward Atmosphere Scale (Third Edition), 14th MMY
Watson-Barker Listening Test, 10th MMY
Watson-Barker Listening Test—High School Version, 11th MMY
Weller-Strawser Scales of Adaptive Behavior for the Learning Disabled, 10th MMY
Work Environment Scale, Second Edition, 12th MMY

7.14.11 Multi-Aptitude Batteries

Armed Services Vocational Aptitude Battery [Forms 18/19], 9th MMY

Comprehensive Ability Battery, 9th MMY
Differential Aptitude Tests, Fifth Edition, 12th MMY
Differential Aptitude Tests—Computerized Adaptive Edition, 12th MMY
Differential Aptitude Tests for Personnel and Career Assessment, 12th MMY
Perceptual-Motor Assessment for Children & Emotional/Behavioral Screening Program, 12th MMY
Quick Cognitive Inventory, 11th MMY
USES General Aptitude Test Battery, 9th MMY
USES Nonreading Aptitude Test Battery, 1982 Edition, 9th MMY

7.14.12 Neuropsychological

Adult Neuropsychological Questionnaire, 12th MMY
Aphasia Diagnostic Profiles, 13th MMY
Apraxia Battery for Adults, 9th MMY
Bedside Evaluation and Screening Test of Aphasia, 11th MMY
Bedside Evaluation Screening Test, Second Edition, 14th MMY
Behavior Change Inventory, 11th MMY
Behavioural Assessment of the Dysexecutive Syndrome, 14th MMY
Booklet Category Test, Second Edition (The), 9th MMY, 14th MMY
Brief Neuropsychological Cognitive Examination, 14th MMY
Brief Test of Head Injury, 13th MMY
Burns Brief Inventory of Communication and Cognition, 14th MMY
California Computerized Assessment Package, 14th MMY
California Verbal Learning Test, Children's Version, 13th MMY
California Verbal Learning Test, Research Edition, Adult Version, 13th MMY
CERAD (Consortium to Establish a Registry for Alzheimer's Disease) Assessment Battery, 12th MMY
Child Neuropsychological Questionnaire, 12th MMY
Children's Auditory Verbal Learning Test-2, 12th MMY
Children's Memory Scale, 14th MMY
Clifton Assessment Procedures for the Elderly, 9th MMY
Clock Test (The), 14th MMY
Closed Head Injury Screener, 14th MMY
Cognistat (The Neurobehavioral Cognitive Status Examination), 14th MMY
Cognitive Behavior Rating Scales, 11th MMY
Cognitive Symptom Checklists, 14th MMY
CogScreen Aeromedical Edition, 14th MMY
Comprehensive Test of Visual Functioning, 12th MMY

Computerized Assessment of Response Bias: Revised Edition, 14th MMY

Consortium to Establish a Registry for Alzheimer's Disease Assessment Battery. *See* CERAD(Consortium to Establish a Registry for Alzheimer's Disease) Assessment Battery

Contextual Memory Test, 14th MMY

Dementia Rating Scale, 11th MMY

Digit Vigilance Test, 14th MMY

Digital Finger Tapping Test, 12th MMY

Discourse Comprehension Test, 13th MMY

Functional Linguistic Communication Inventory, 13th MMY

Halstead Russell Neuropsychological Evaluation System, 12th MMY

Halstead-Reitan Neuropsychological Test Battery, 9th MMY

Human Information Processing Survey, 10th MMY

Infanib (The), 14th MMY

Intermediate Booklet Category Test, 14th MMY

Kaufman Short Neuropsychological Assessment Procedure, 13th MMY

Learning and Memory Battery, 14th MMY

Learning Disabilities Diagnostic Inventory, 14th MMY

Loewenstein Occupational Therapy Cognitive Assessment, 13th MMY

Luria-Nebraska Neuropsychological Battery: Children's Revision, 11th MMY

Luria-Nebraska Neuropsychological Battery: Forms I and II, 11th MMY

McCarron-Dial System, 11th MMY

MicroCog: Assessment of Cognitive Functioning, 13th MMY

Mini Inventory of Right Brain Injury, 11th MMY

Minnesota Test for Differential Diagnosis of Aphasia, 13th MMY

National Adult Reading Test, Second Edition, 12th MMY

NEPSY: A Developmental Neuropsychological Assessment, 14th MMY

Neurobehavioral Cognitive Status Examination. *See* Cognistat (The Neurobehavioral Cognitive Status Examination)

Neurological Dysfunctions of Children, 9th MMY

Neuropsychological Impairment Scale (The), 14th MMY

Neuropsychological Questionnaire, 9th MMY

Neuropsychology Behavior and Affect Profile, 14th MMY

Ohio Functional Assessment Battery: Standardized Tests for Leisure and Living Skills, 14th MMY

Philadelphia Head Injury Questionnaire, 12th MMY

Portable Tactual Performance Test, 11th MMY

Portland Digit Recognition Test, 12th MMY

Quick Neurological Screening Test, 2nd Revised Edition, 14th MMY
Repeatable Battery for the Assessment of Neuropsychological Status, 14th MMY
Rey Auditory Verbal Learning Test: A Handbook, 14th MMY
Rey Complex Figure Test and Recognition Trial, 14th MMY
Rivermead Behavioural Memory Test [Second Edition] (The), 14th MMY
Rivermead Perceptual Assessment Battery, 11th MMY
Ross Information Processing Assessment, Second Edition, 14th MMY
Ross Information Processing Assessment—Geriatric, 14th MMY
Scales of Adult Independence, Language and Recall, 14th MMY
Scales of Cognitive Ability for Traumatic Brain Injury, 13th MMY
Screening Test for the Luria-Nebraska Neuropsychological Battery: Adult and Children's Forms, 11th MMY
Sensory Integration and Praxis Tests, 12th MMY
Severe Cognitive Impairment Profile, 14th MMY
Short Category Test, Booklet Format, 11th MMY
Sklar Aphasia Scale, Revised 1983, 10th MMY
Stroop Color and Word Test, 9th MMY
Stroop Neuropsychological Screening Test, 11th MMY
Test of Memory Malingering, 14th MMY
Test of Oral and Limb Apraxia, 13th MMY
Test of Orientation for Rehabilitation Patients, 13th MMY
Visual Search and Attention Test, 12th MMY
WAIS-R NI, 14th MMY
Word Memory Test (The), 14th MMY

7.14.13 Personality

Abuse Risk Inventory for Women, Experimental Edition, 11th MMY
ACDI-Corrections Version and Corrections Version II, 14th MMY
Achievement Motivation Profile, 14th MMY
Additional Personality Factor Inventory—2, 9th MMY
Adjective Check List (The), 9th MMY
Adolescent Apperception Cards, 14th MMY
Adolescent Coping Scale, 13th MMY
Adolescent Dissociative Experiences Scale, 14th MMY
Adolescent Drinking Index, 12th MMY
Adolescent Psychopathology Scale, 14th MMY
Adolescent Separation Anxiety Test, 9th MMY
Adult Personality Inventory, 12th MMY

Adult Personality Inventory [Revised], 14th MMY
Adult Self Expression Scale, 9th MMY
Adult Suicidal Ideation Questionnaire, 12th MMY
Affective Perception Inventory, 9th MMY
Affective Perception Inventory/College Level (The), 11th MMY
Age Projection Test, 12th MMY
Alcadd Test, Revised Edition (The), 11th MMY
Alcohol Dependence Scale, 10th MMY
Alcohol Use Inventory, 12th MMY
Anomalous Sentences Repetition Test (The), 12th MMY
APT Inventory (The), 13th MMY
Arousal Seeking Tendency Scale, 13th MMY
Assessment of Core Goals, 12th MMY
Assessment of Qualitative and Structural Dimensions of Object Representations, Revised Edition,9th MMY
Athletic Motivation Inventory, 12th MMY
Attentional and Interpersonal Style Inventory (The), 14th MMY
Attitudes Toward Mainstreaming Scale, 12th MMY
Attitudes Toward Working Mothers Scale, 9th MMY
Balanced Emotional Empathy Scale (The), 13th MMY
Barclay Classroom Assessment System (The), 9th MMY
BarOn Emotional Quotient Inventory, 14th MMY
Basic Living Skills Scale, 10th MMY
Basic Personality Inventory, 12th MMY
BASIS-A Inventory [Basic Adlerian Skills for Interpersonal Success—Adult Form], 13th MMY
Battery for Health Improvement, 14th MMY
Beck Anxiety Inventory [1993 Edition], 13th MMY
Beck Depression Inventory—Ⅱ, 13th MMY, 14th MMY
Beck Hopelessness Scale [Revised], 13th MMY
Beck Scale for Suicide Ideation, 13th MMY
Behavior Rating Scale, 9th MMY
Behavioral Academic Self-Esteem, 9th MMY
Behaviour Problems: A System of Management, 11th MMY
Bell Object Relations and Reality Testing Inventory, 14th MMY
Bem Sex-Role Inventory, 9th MMY
Bessell Measurement of Emotional Maturity Scales, 9th MMY
BEST Instruments, 13th MMY

Biographical and Personality Inventory, Series Ⅱ, 11th MMY
Biographical Inventory Form U, 9th MMY
Bloom Sentence Completion Survey, 9th MMY
BrainMap (The), 11th MMY
Butcher Treatment Planning Inventory, 14th MMY
California Child Q-Set (The), 9th MMY
California Psychological Inventory, Revised Edition, 11th MMY
California Q-Sort (Revised Adult Set), 14th MMY
California Test of Personality, 9th MMY
Callahan Anxiety Pictures, 9th MMY
Campbell Leadership Index, 12th MMY
Canfield Instructional Styles Inventory, 11th MMY
Canfield Learning Styles Inventory, 11th MMY
Career Beliefs Inventory, 12th MMY
Carey Temperament Scales, 14th MMY
Carlson Psychological Survey, 9th MMY
Carroll Depression Scales, 14th MMY
Chart of Initiative and Independence, 9th MMY
Charteris Reading Test, 12th MMY
Child Abuse Potential Inventory, Form VI (The), 10th MMY
Child and Adolescent Adjustment Profile, 9th MMY
Child Anxiety Scale, 9th MMY
Childhood Autism Rating Scale (The), 11th MMY
Children of Alcoholics Screening Test, 10th MMY
Children's Academic Intrinsic Motivation Inventory, 10th MMY
Children's Apperception Test [1991 Revision], 13th MMY
Children's Apperceptive Story-Telling Test, 12th MMY
Children's Depression Inventory, 11th MMY
Children's Depression Rating Scale, Revised, 14th MMY
Children's Depression Scale [Second Research Edition], 12th MMY
Children's Inventory of Self-Esteem, 12th MMY
Children's Personality Questionnaire, 1985 Edition, 13th MMY
Children's Problems Checklist, 10th MMY
Children's Role Inventory, 13th MMY
Christensen Dietary Distress Inventory, 13th MMY
Clinical Analysis Questionnaire, 9th MMY
Clyde Mood Scale, 9th MMY
College Adjustment Scales, 13th MMY

Communication Response Style: Assessment, 11th MMY
Composite International Diagnostic Interview, 13th MMY
Comprehensive Assessment of Symptoms and History, 12th MMY
Comprehensive Personality Profile, 14th MMY
Conflict Management Appraisal, 12th MMY
Conflict Management Survey, 12th MMY
Conflict Style Inventory, 14th MMY
Coolidge Axis II Inventory, 14th MMY
Coopersmith Self-Esteem Inventories, 9th MMY
Coping Inventory for Stressful Situations, 14th MMY
Coping Resources Inventory, 12th MMY
Coping Responses Inventory—Adult and Youth, 14th MMY
Coping Resources Inventory for Stress, 13th MMY
Coping with Stress, 12th MMY
Couples BrainMap (The), 11th MMY
Couple's Pre-Counseling Inventory, Revised Edition, 1987, 11th MMY
CPF [Second Edition], 12th MMY
Creative Behavior Inventory, 11th MMY
Creative Reasoning Test (The), 11th MMY
Creatrix Inventory (The), 11th MMY
Cross-Cultural Adaptability Inventory, 14th MMY
Crown-Crisp Experiential Index, 9th MMY
Culture-Free Self-Esteem Inventories, Second Edition, 12th MMY
Custody Quotient (The), 11th MMY
Daily Stress Inventory, 11th MMY
Dating Problems Checklist, 9th MMY
Davidson Trauma Scale, 14th MMY
Decision Making Inventory, 10th MMY
Defense Mechanisms Inventory [Revised], 13th MMY
Defining Issues Test, 11th MMY
Denver Community Mental Health Questionnaire—Revised, 9th MMY
Depression and Anxiety in Youth Scale, 13th MMY
Derogatis Affects Balance Scale [Revised], 14th MMY
Derogatis Psychiatric Rating Scale, 13th MMY
Derogatis Sexual Functioning Inventory, 9th MMY
Derogatis Stress Profile, 13th MMY
Description of Body Scale, 9th MMY
Devereux Scales of Mental Disorders, 14th MMY

Diagnostic Interview for Borderline Patients, 10th MMY
Diagnostic Screening Batteries (The), 13th MMY
Differential Diagnostic Technique (The), 14th MMY
Dimensions of Self-Concept, 11th MMY
Dissociative Experiences Scale, 12th MMY
Diversity Awareness Profile, 12th MMY
Domestic Violence Inventory, 14th MMY
Draw a Person: Screening Procedure for Emotional Disturbance, 12th MMY
Draw-a-Story: Screening for Depression and Age or Gender Differences [Revised], 13th MMY
Dyadic Adjustment Scale, 11th MMY
E. S. Survey, 9th MMY
Early Childhood Inventory, 14th MMY
Early Coping Inventory, 11th MMY
Early Memories Procedure, 13th MMY
Eating Disorder Inventory-2, 12th MMY
Edinburgh Questionnaires (The) (1982 Edition), 10th MMY
Ego Function Assessment, 13th MMY
EMO Questionnaire, 9th MMY
Emotional Behavioral Checklist, 11th MMY
Emotional Problems Scales, 13th MMY
Employee Assistance Program Inventory, 14th MMY
Employee Reliability Inventory, 12th MMY
Endler Multidimensional Anxiety Scales, 12th MMY
EPS Sentence Completion Technique, 13th MMY
Executive Profile Survey, 12th MMY
Facial Interpersonal Perception Inventory (The), 9th MMY
Family Apperception Test, 12th MMY
Family Relations Test: Children's Version, 11th MMY
Family Relationship Inventory, 10th MMY
Feelings, Attitudes, and Behaviors Scale for Children, 14th MMY
Firestone Assessment of Self-Destructive Thoughts, 14th MMY
FIRO Awareness Scales (The), 9th MMY
Friedman Well-Being Scale, 14th MMY
General Health Questionnaire, 12th MMY
Global Assessment Scale, 11th MMY
Golombok Rust Inventory of Sexual Satisfaction (The), 10th MMY
Gordon Personal Profile—Inventory, 13th MMY

Grief Experience Inventory (The), 12th MMY
Group Achievement Identification Measure, 11th MMY
Group Inventory for Finding Creative Talent, 9th MMY
Group Styles Inventory, 12th MMY
Guilford-Zimmerman Temperament Survey (The), 9th MMY
Hall-Tonna Inventory of Values (The), 11th MMY
Hamilton Depression Inventory, 13th MMY
Hand Test [Revised] (The), 10th MMY, 14th MMY
Happiness Measures, 11th MMY
Harding Stress-Fair Compatibility Test, 9th MMY
Hare Psychopathy Checklist: Screening Version, 14th MMY
Hare Psychopathy Checklist—Revised (The), 12th MMY
Hassles and Uplifts Scales, Research Edition, 11th MMY
Health Attribution Test, 11th MMY
High School Personality Questionnaire, 11th MMY
Hilson Adolescent Profile, 11th MMY
Hogan Development Survey, 14th MMY
Hogan Personality Inventory [Revised], 13th MMY
Holden Psychological Screening Inventory, 14th MMY
Holtzman Inkblot Technique, 9th MMY
House-Tree-Person and Draw-a-Person as Measures of Abuse in Children: A Quantitative Scoring System, 13th MMY
Howarth Mood Adjective Check List, 9th MMY
Howarth Personality Questionnaire, 9th MMY
Human Figure Drawing Test, 14th MMY
Human Resource Development Report, 11th MMY
Illness Behavior Questionnaire, Second Edition, 10th MMY
Imagery and Disease, 12th MMY
Impact Message Inventory, Research Edition, 12th MMY
Independent Living Behavior Checklist, 9th MMY
Index of Personality Characteristics, 11th MMY
Individual Style Survey, 13th MMY
Inquiry Mode Questionnaire: A Measure of How You Think and Make Decisions, 12th MMY
INSIGHT Inventory, 13th MMY
INSIGHT Inventory [Revised], 14th MMY
Interaction Index, 13th MMY
Internalizing Symptoms Scale for Children, 14th MMY

Interpersonal Adjective Scales, 14th MMY
Interpersonal Behavior Survey, 9th MMY
Interpersonal Style Inventory, 10th MMY
Intrex Questionnaires, 12th MMY
Inventory of Positive Thinking Traits (The), 13th MMY
Inventory of Suicide Orientation-30, 13th MMY
Inwald Personality Inventory [Revised], 12th MMY
Iowa Parent Behavior Inventory, 9th MMY
Jackson Personality Inventory—Revised, 13th MMY
Jenkins Activity Survey, 9th MMY
Jesness Inventory (The), 14th MMY
Job Attitude Scale, 11th MMY
Joseph Pre-School and Primary Self-Concept Screening Test, 9th MMY
Juvenile Automated Substance Abuse Evaluation, 14th MMY
Keegan Type Indicator, 10th MMY
Khatena Torrance Creative Perception Inventory, 13th MMY
Kinetic Drawing System for Family and School: A Handbook, 10th MMY
Kirton Adaption-Innovation Inventory, 11th MMY
Kirton Adaption-Innovation Inventory [1998 Edition], 14th MMY
Kohn Problem Checklist/Kohn Social Competence Scale, 13th MMY
Kolbe Conative Index, 12th MMY, 14th MMY
La Monica Empathy Profile, 10th MMY
Lake St. Clair Incident (The), 9th MMY
Leadership and Self-Development Scale, 9th MMY
Leadership Competency Inventory, 13th MMY
Leadership Skills Inventory, 10th MMY
Learning Channel Preference Checklist (The), 11th MMY
Learning Preference Inventory, 11th MMY
Learning Process Questionnaire, 11th MMY
Learning Style Identification Scale, 9th MMY
Learning Style Profile, 12th MMY
Learning Styles Inventory [Creative Learning Press, Inc.], 9th MMY
Leeds Scales for the Self-Assessment of Anxiety and Depression, 9th MMY
Leisure Diagnostic Battery (The), 12th MMY
Let's Talk Inventory for Children, 11th MMY
Lewis Counseling Inventory, 9th MMY
Life Orientation Inventory, 11th MMY
Life Roles Inventory, 13th MMY

Life Space Analysis Profile, 13th MMY
Life Styles Inventory, 12th MMY
London House Personnel Selection Inventory, 9th MMY
Maferr Inventory of Feminine Values, 9th MMY
Maferr Inventory of Masculine Values, 9th MMY
Malingering Probability Scale, 14th MMY
Management Style Inventory [Training House, Inc.], 11th MMY
Manchester Personality Questionnaire, 13th MMY
Manson Evaluation, Revised Edition (The), 11th MMY
Marital Communication Inventory (A), 9th MMY
Marital Satisfaction Inventory—Revised, 9th MMY, 14th MMY
Martinek-Zaichkowsky Self-Concept Scale for Children, 9th MMY
Maryland Addictions Questionnaire, 14th MMY
Maslach Burnout Inventory, 10th MMY
mCircle Instrument (The), 11th MMY
Meadow-Kendall Social-Emotional Assessment Inventory for Deaf and Hearing Students,10th MMY
Measure of Achieving Tendency, 13th MMY
Measure of Child Stimulus Screening and Arousability, 9th MMY
Measures of Affiliative Tendency and Sensitivity to Rejection, 13th MMY
Measures of Pleasure-, Arousal-, and Dominance-Inducing Qualities of Parental Attitudes,9th MMY
Measures of Psychosocial Development, 11th MMY
Meeker Behavioral Correlates, 10th MMY
Memory and Behavior Problems Checklist and The Burden Interview, 14th MMY
Menstrual Distress Questionnaire, 12th MMY
Mental Status Checklist for Adolescents, 11th MMY
Mental Status Checklist for Children, 11th MMY
Meta-Motivation Inventory, 9th MMY
Meyer-Kendall Assessment Survey, 12th MMY
Miller Motivation Scale, 10th MMY
Milon Adolescent Clinical Inventory, 12th MMY
Millon Clinical Multiaxial Inventory—III [Manual Second Edition], 14th MMY
Millon Clinical Multiaxial Inventory—III [Second Edition], 13th MMY
Millon Index of Personality Styles, 13th MMY
MindMaker6, 11th MMY
Miner Sentence Completion Scale, 11th MMY
Minnesota Multiphasic Personality Inventory-Adolescent, 12th MMY

Minnesota Multiphasic Personality Inventory-2, 11th MMY
Mother-Child Relationship Evaluation, 1980 Edition (The), 12th MMY
Motivation Analysis Test, 12th MMY
Motivational Patterns Inventory, 12th MMY
Motives, Values, Preferences Inventory, 14th MMY
Multidimensional Addictions and Personality Profile (The), 14th MMY
Multidimensional Anxiety Scale for Children, 14th MMY
Multidimensional Assessment and Planning Form, 9th MMY
Multidimensional Self Concept Scale, 12th MMY
Multidimensional Self-Esteem Inventory (The), 11th MMY
Multiple Affect Adjective Check List, Revised, 10th MMY
Multiscore Depression Inventory, 11th MMY
Multiscore Depression Inventory for Children, 14th MMY
Murphy-Meisgeier Type Indicator for Children, 12th MMY
Mutually Responsible Facilitation Inventory, 9th MMY
Myers-Briggs Type Indicator, 10th MMY
Myers-Briggs Type Indicator, Form M, 14th MMY
NEEDS Survey, 14th MMY
NEO-4, 14th MMY
North American Depression Inventories for Children and Adults, 11th MMY
NPF [Second Edition], 12th MMY
Observational Emotional Inventory—Revised, 11th MMY
Occupational Personality Questionnaire, 11th MMY
Occupational Stress Indicator, 11th MMY
Occupational Stress Inventory, 11th MMY
Occupational Stress Inventory—Revised Edition, 14th MMY
Occupational Type Profile, 14th MMY
Oetting's Computer Anxiety Scale, 10th MMY
Offer Self-Image Questionnaire, Revised, 12th MMY
Oliver Organizatioin Desscription, Questionnaire, 11th MMY
Pain Patient Profile, 14th MMY
P. A. R. Admissions Test, 11th MMY
Parent as a Teacher Inventory [Revised], 13th MMY
Parent Awareness Skills Survey, 12th MMY
Parent-Child Relationship Inventory, 13th MMY
Parenting Stress Index, Third Edition, 13th MMY
Partner Relationship Inventory (Research Edition), 12th MMY
Perception of Ability Scale for Students, 12th MMY

Perception-of-Relationships-Test, 12th MMY
Personal Experience Inventory, 11th MMY
Personal History Checklist for Adults, 11th MMY
Personal Inventory of Needs, 12th MMY
Personal Orientation Dimensions, 14th MMY
Personal Problems Checklist—Adult, 10th MMY
Personal Problems Checklist for Adolescents, 10th MMY
Personal Profile System, 10th MMY
Personal Resource Questionnaire, 11th MMY
Personal Skills Map, 13th MMY
Personal Stress Assessment Inventory, 14th MMY
Personal Style Assessment, 11th MMY
Personal Style Assessment, Jung-Parry Form, 13th MMY
Personal Style Indicator, 13th MMY
Personal Style Inventory, 14th MMY
Personal Styles Inventory [PSI-120], 13th MMY
Personal Values Questionnaire, 13th MMY
Personality Adjective Check List, 12th MMY
Personality Assessment Inventory, 12th MMY
Personality Assessment Screener, 14th MMY
Personality Disorder Interview-IV: A Semistructured Interview for the Assessment of Personality Disorders, 14th MMY
Personality Inventory for Children, Revised Format, 10th MMY
Personality Inventory for Youth, 13th MMY
Personality Research Form, 3rd Edition, 10th MMY
PERSONALYSIS, 12th MMY
Phase II Profile Integrity Status Inventory and Addendum, 10th MMY
Piers-Harris Children's Self-Concept Scale (The Way I Feel About Myself), 9th MMY
Porteous Problem Checklist, 10th MMY
Positive and Negative Syndrome Scale, 12th MMY
Posttraumatic Stress Diagnostic Scale, 14th MMY
Problem Experiences Checklist, 12th MMY
Problem Solving Inventory (The), 11th MMY
Problem-Solving Decision-Making Style Inventory, 10th MMY
Prout-Strohmer Personality Inventory, 11th MMY
Psychap Inventory (The), 11th MMY
Psychiatric Content Analysis and Diagnosis, 14th MMY

Psychiatric Diagnostic Interview—Revised, 11th MMY
Psychosocial Adjustment to Illness Scale, 10th MMY
Psychotherapy Outcome Kit (Including Quality of Emotional Life Self-Report), 14th MMY
Quality of Life Enjoyment and Satisfaction Questionnaire, 14th MMY
Quality of Life Questionnaire, 11th MMY
Questionnaire Measure of Individual Differences in Achieving Tendency (A), 10th MMY
Questionnaire Measure of Trait Arousability (Or Its Converse, Stimulus Screening), 13th MMY
Questionnaire on Resources and Stress, 11th MMY
Racial Attitude Test, 11th MMY
Reid Report, 12th MMY
Reiss-Epstein-Gursky Anxiety Sensitivity Index, 11th MMY
Revised Children's Manifest Anxiety Scale, 10th MMY
Revised Hamilton Rating Scale for Depression, 13th MMY
Revised NEO Personality Inventory me, 12th MMY
Reynolds Adolescent Depression Scale, 11th MMY
Reynolds Child Depression Scale, 11th MMY
Reynolds Depression Screening Inventory, 14th MMY
Roberts Apperception Test for Children [with new Interpretive Handbook], 14th MMY
Rokeach Value Survey, 12th MMY
Rorschach, 14th MMY
Rotter Incomplete Sentences Blank, Second Edition, 12th MMY
Rust Inventory of Schizotypal Cognitions, 13th MMY
Sales Personality Questionnaire, 12th MMY
Sales Transaction Audit, 14th MMY
Salience Inventory (The), 11th MMY
SAQ-Adult Probation III, 14th MMY
Scale for Assessing Emotional Disturbance, 14th MMY
Scale for the Assessment of Negative Symptoms, 12th MMY
Scale for the Assessment of Positive Symptoms, 12th MMY
Scale of Social Development, 11th MMY
Schedule for Affective Disorders and Schizophrenia, Third Edition, 12th MMY
Schedule for Nonadaptive and Adaptive Personality, 14th MMY
Schedule of Recent Experience (The), 10th MMY
School Attitude Measure, Second Edition, 11th MMY
Search Institute Profiles of Student Life, 11th MMY

Self Worth Inventory, 13th MMY
Self-Awareness Profile, 11th MMY
Self-Description Questionnaire—I, II, III, 13th MMY
Self-Esteem Index, 11th MMY
Self-Perception Inventory: Nursing Forms, 11th MMY
Self-Perception Profile for College Students, 11th MMY
Senior Apperception Technique [1985 Revision] (The), 13th MMY
Sentence Completion Series, 14th MMY
Severity and Acuity of Psychiatric Illness Scales—Adult Version (The), 14th MMY
Sexual Adjustment Inventory, 14th MMY
Shapiro Control Inventory, 13th MMY
Singer-Loomis Inventory of Personality, Experimental Edition (The), 10th MMY
Singer-Loomis Type Deployment Inventory, 14th MMY
SIPOAS: Styles in Perception of Affect Scale, 14th MMY
Sixteen Personality Factor Questionnaire, Fifth Edition, 12th MMY
Social Phobia and Anxiety Inventory, 14th MMY
Social Reticence Scale, 11th MMY
Social Skills Inventory, Research Edition, 11th MMY
Social Styles Analysis, 11th MMY
Social-Emotional Dimension Scale, 11th MMY
SPECTRUM-I: A Test of Adult Work Motivation, 12th MMY
Spiritual Well-Being Scale, 12th MMY
Staff Burnout Scale for Health Professionals, 9th MMY
Stanton Inventory (The), 9th MMY
Stanton Survey and the Stanton Survey Phase II (The), 9th MMY
State Trait-Depression Adjective Check Lists, 13th MMY
State-Trait Anger Expression Inventory [1996 Edition], 13th MMY, 14th MMY
Stokes/Gordon Stress Scale, 13th MMY
Strength Deployment Inventory, 14th MMY
Stress Analysis System, 10th MMY
Stress Audit, 10th MMY
Stress Impact Scale, 12th MMY
Stress Index for Parents of Adolescents, 14th MMY
Stress Management Questionnaire, 10th MMY
Stress Resiliency Profile, 13th MMY
Stress Response Scale, 11th MMY
Structured Clinical Interview for DSM-IV Axis I Disorders: Clinician Version, 14th MMY

Structured Clinical Interview for DSM-IV Axis II Personality Disorders, 14th MMY
Structured Interview of Reported Symptoms, 12th MMY
Structured Pediatric Psychosocial Interview, 10th MMY
Student Adaptation to College Questionnaire, 11th MMY
Student Adjustment Inventory, 12th MMY
Student Developmental Task and Lifestyle Inventory, 11th MMY
Student Referral Checklist, 11th MMY
Student Self-Concept Scale, 13th MMY
Student Styles Questionnaire, 14th MMY
Student Talent and Risk Profile, 12th MMY
Study Process Questionnaire, 11th MMY
Style of Learning and Thinking, 11th MMY
Substance Abuse Life Circumstance Evaluation, 14th MMY
Substance Abuse Subtle Screening Inventory, 12th MMY
Suicidal Ideation Questionnaire, 11th MMY
SUPER Star Profiles, 13th MMY
Supervisory Human Relations, 11th MMY
Survey of Organizational Climate, 11th MMY
Survey of Organizational Culture, 11th MMY
Survey of Personal Values, 10th MMY
Survey of Work Values, Revised, Form U, 12th MMY
Symptom Assessment-45 Questionnaire, 14th MMY
Symptom Checklist-90-Revised, 9th MMY
Symptom Scale—77, 14th MMY
Taylor-Johnson Temperament Analysis [1992 Edition], 13th MMY, 14th MMY
Teacher Attitude Inventory: Identifying Teacher Positions in Relation to Educational Issues and Decisions (A), 10th MMY
Teacher Stress Inventory, 11th MMY
Team Development Survey, 13th MMY
TEMAS (Tell-Me-A-Story), 11th MMY
Temperament Assessment Battery for Children, 11th MMY
Tennessee Self-Concept Scale, Second Edition, 13th MMY
Test Anxiety Profile, 9th MMY
Test Attitude Inventory, 9th MMY
Test for Creative Thinking—Drawing Production, 14th MMY
Themes Concerning Blacks, 9th MMY
Thomas-Kilmann Conflict Mode Instrument, 10th MMY
Tiffany Control Scales, 14th MMY

Time Perception Inventory, 11th MMY
TMJ Scale, 12th MMY
Trait Pleasure-Displeasure Scale (The), 13th MMY
Trauma Symptom Inventory, 14th MMY
Treatment Intervention Inventory, 14th MMY
Understanding and Managing Stress, 12th MMY
Validity Indicator Profile, 14th MMY
Values Scale, Second Edition (The), 13th MMY
Victoria Symptom Validity Test, 14th MMY
VITAL Checklist and Curriculum Guide, 11th MMY
Waksman Social Skills Rating Scale, 10th MMY
Walker-McConnell Scale of Social Competencc and School Adjustment, 11th MMY
Ways of Coping Questionnaire, Research Edition, 11th MMY
Weinberg Depression Scale for Children and Adolescents, 14th MMY
What Do You Say?,12th MMY
Whitaker Index of Schizophrenic Thinking, 11th MMY
Wide Range Interest-Opinion Test, 9th MMY
Work Personality Profile, 11th MMY

7.14.14 Reading

ACER Applied Reading Test, 12th MMY
Assessing Reading Difficulties: A Diagnostic and Remedial Approach, 9th MMY
Bader Reading and Language Inventory, 10th MMY
Basic Reading Inventory, Seventh Edition, 12th MMY, 14th MMY
Bench Mark Measures, 9th MMY
Boder Test of Reading-Spelling Patterns (The), 9th MMY
Burns/Roe Informal Reading Inventory: Preprimer to Twelfth Grade, Third Edition, 12th MMY
Burns/Roe Informal Reading Inventory: Preprimer to Twelfth Grade, Fifth Edition, 14th MMY
Burt Word Reading Test, New Zealand Revision, 9th MMY
California Diagnostic Reading Tests, 11th MMY
CHILD Center Operational Assessment Tool (The), 9th MMY
Classroom Reading Inventory, Seventh Edition, 10th MMY
Cloze Reading Tests 1-3, Second Edition, 9th MMY
Clymer-Barrett Readiness Test, Revised Edition, 9th MMY
Content Inventories: English, Social Studies, Science, 9th MMY
Corrective Reading Mastery Tests, 9th MMY

Croft Readiness Assessment in Comprehension Kit, 9th MMY
Decoding Skills Test, 10th MMY
Degrees of Reading Power, 12th MMY
Degrees of Reading Power [Revised], 14th MMY
Degrees of Word Meaning, 13th MMY
Diagnostic Assessments of Reading, 12th MMY
Diagnostic Reading Scales, 9th MMY
Diagnostic Screening Test: Reading, Third Edition, 9th MMY
Durrell Analysis of Reading Difficulty, Third Edition, 9th MMY
Dyslexia Screening Instrument, 14th MMY
Dyslexia Screening Survey (The), 10th MMY
Edinburgh Reading Tests, 9th MMY
Ekwall/Shanker Reading Inventory—Third Edition, 13th MMY
GAP Reading Comprehension Test, Third Edition, 10th MMY
Gates-MacGinitie Reading Tests, Canadian Edition, 9th MMY
Gates-MacGinitie Reading Tests, Third Edition, 11th MMY
Gates-McKillop-Horowitz Reading Diagnostic Test, Second Edition, 9th MMY
Gray Oral Reading Tests, Third Edition, 12th MMY
Gray Oral Reading Tests—Diagnostic, 11th MMY
Group Diagnostic Reading Aptitude and Achievement Tests, Intermediate Form, 12th MMY
Group Reading Test, Third Edition, 12th MMY
Group Reading Test, Fourth Edition, 14th MMY
Industrial Reading Test-Infant Reading Tests (The), 9th MMY
Informal Reading Comprehension Placement Test[Revised], 13th MMY
IOX Basic Skills Word List (The), 9th MMY
Johnston Informal Reading Inventory, 12th MMY
Laurita-Trembley Diagnostic Word Processing Test (The), 9th MMY
Lollipop Test: A Diagnostic Screening Test of School Readiness—Revised (The), 11th MMY
London Reading Test, 9th MMY
Macmillan Diagnostic Reading Pack (The), 9th MMY
Macmillan Graded Word Reading Test, 10th MMY
Macmillan Group Reading Test, 10th MMY
Maryland/Baltimore County Design for Adult Basic Education, 9th MMY
McCall-Crabbs Standard Test Lessons in Reading, 9th MMY
McCarthy Individualized Diagnostic Reading Inventory, Revised Edition, 9th MMY
Murphy-Durrell Reading Readiness Screen, 11th MMY

Neale Analysis of Reading Ability, Revised British Edition, 11th MMY
Nelson Reading Skills Test, Forms 3 and 4 (The), 9th MMY
Nelson-Denny Reading Test, Forms G and H, 13th MMY
New Macmillan Reading Analysis, 10th MMY
NewGAP, 12th MMY
Progressive Achievement Tests in Reading: Reading Comprehension and Reading Vocabulary Tests, 11th MMY
Progressive Achievement Tests of Reading [Revised], 12th MMY
Ramsay Corporation Job Skills—Health Care Reading Test, 13th MMY
Ramsay Corporation Job Skills—Office Reading Test, 13th MMY
Ramsay Corporation Job Skills—Supervisory Reading Test, 13th MMY
Reading Ability Series, 11th MMY
Reading Comprehension Battery for Aphasia, Second Edition, 14th MMY
Reading Comprehension Inventory, 11th MMY
Reading Comprehension Test DE, 10th MMY
Reading Evaluation Adult Diagnosis (Revised), 11th MMY
Reading Evaluation Adult Diagnosis, Fifth Edition, 14th MMY
Reading Evaluation and Diagnostic Screen, 10th MMY
Reading Style Inventory, 11th MMY
Reading Test AD, 10th MMY
Reading Test SR-A and SR-B, 10th MMY
Reversals Frequency Test (The), 11th MMY
Revised PSB—Reading Comprehension Examination, 13th MMY
Roswell-Chall Diagnostic Reading Test of Word Analysis Skills, Revised and Extended, 14th MMY
School Readiness Test, 12th MMY
Secondary & College Reading Inventory, Second Edition, 13th MMY
Shortened Edinburgh Reading Test, 14th MMY
Slosson Oral Reading Test [Revised], 12th MMY
Slosson Test of Reading Readiness, 12th MMY
SPAR Spelling and Reading Tests, Second Edition, 12th MMY
SPAR Spelling and Reading Tests, Third Edition, 14th MMY
Standardized Reading Inventory, 10th MMY
Standardized Reading Inventory, Second Edition, 14th MMY
Stanford Diagnostic Reading Test, Fourth Edition, 13th MMY
STAR Reading, 14th MMY
Stieglitz Informal Reading Inventory: Assessing Reading Behaviors from Emergent to Advanced Levels (The), 13th MMY

Suffolk Reading Scale, 11th MMY
Test of Awareness of Language Segments, 11th MMY
Test of Early Reading Ability—2, 11th MMY
Test of Early Reading Ability—Deaf or Hard of Hearing, 12th MMY
Test of Inference Ability in Reading Comprehension, 13th MMY
Test of Initial Literacy, 11th MMY
Test of Reading Comprehension, Third Edition, 13th MMY
Tests of Reading Comprehension, 11th MMY
Watson-Glaser Critical Thinking Appraisal, 9th MMY
Wide-Span Reading Test, 10th MMY
Woodcock Diagnostic Reading Battery, 14th MMY
Woodcock Reading Mastery Tests—Revised [1998 Normative Update], 14th MMY
Woodcock Reading Mastery Tests—Revised [1998 Norms], 10th MMY
Word Recognition and Phonic Skills, 14th MMY

7.14.15 Science

Scientific Orientation Test, 14th MMY
Tapping Students' Science Beliefs: A Resource for Teaching and Learning, 13th MMY

7.14.16 Sensorimotor

Basic Visual-Motor Association Test, 9th MMY
Bender-Gestalt Test, 11th MMY
Benton Visual Retention Test, Fifth Edition, 13th MMY
Bruininks-Oseretsky Test of Motor Proficiency, 9th MMY
Carrow Auditory-Visual Abilities Test, 9th MMY
City University Colour Vision Test, Second Edition 1980 (The), 13th MMY
DeGangi-Berk Test of Sensory Integration, 10th MMY
Developmental Test of Visual Perception, Second Edition, 12th MMY
Developmental Test of Visual-Motor Integration [Third Revision], 12th MMY
Developmental Test of Visual-Motor Integration, 4th Edition, Revised, 14th MMY
Dysphagia Evaluation Protocol, 14th MMY
Erhatdt Developmental Vision Assessment, 12th MMY
Evaluating Movement and Posture Disorganization in Dyspraxic Children, 12th MMY
Gibson Spiral Maze, Second Edition, 14th MMY
Grooved Pegboard Test, 12th MMY
Hill Performance Test of Selected Positional Concepts, 9th MMY
Hooper Visual Organization Test (The), 12th MMY

Infant/Toddler Symptom Checklist, 14th MMY
Inventory of Perceptual Skills, 12th MMY
Kent Visual Perception Test, 14th MMY
Lateral Preference Schedule, 11th MMY
McDowell Vision Screening Kit, 14th MMY
McGill Pain Questionnaire (The), 14th MMY
Minnesota Manual Dexterity Test, 12th MMY
Modified Version of the Bender-Gestalt Test for Preschool and Primary School Children (The), 11th MMY
Motor Skills Inventory, 11th MMY
Motor-Free Visual Perception Test—Revised, 14th MMY
Motor-Free Visual Perception Test—Vertical, 14th MMY
OSOT Perceptual Evaluation, 12th MMY
Pin Test (The), 11th MMY
Sensory Evaluation Kit, 13th MMY
Sensory Integration Inventory—Revised for Individuals with Developmental Disabilities, 14th MMY
Test of Visual-Motor Integration, 14th MMY
Test of Visual-Motor Skills, 13th MMY
Wide Range Assessment of Visual Motor Abilities, 14th MMY

7.14.17 Social Studies

Basic Economics Test, 11th MMY
Cultural Literacy Test, 12th MMY
Test of Economic Knowledge, 11th MMY
Test of Economic Literacy, Second Edition, 12th MMY
Test of Understanding in College Economics, Third Edition, 13th MMY

7.14.18 Speech and Hearing

Adapted Sequenced Inventory of Communication Development, 12th MMY
Apraxia Profile: A Descriptive Assessment Tool for Children (The), 14th MMY
Arizona Articulation Proficiency Scale, Second Edition, 11th MMY
Assessing and Teaching Phonological Knowledge, 14th MMY
Assessing Linguistic Behaviors: Assessing Prelinguistic and Early Linguistic Behaviors in Developmentally Young Children, 11th MMY
Assessment of Aphasia and Related Disorders, Second Edition (The), 10th MMY
Assessment of Fluency in School-Age Children, 10th MMY
Assessment of Intelligibility of Dysarthric Speech, 10th MMY

Assessment of Phonological Processes—Revised (The), 12th MMY
Auditory Continuous Performance Test, 13th MMY
Auditory Discrimination and Attention Test (The), 12th MMY
Bankson-Bernthal Test of Phonology, 12th MMY
Bedside Evaluation of Dysphagia, 14th MMY
Behavior Analysis Language Instrument, 11th MMY
Boston Assessment of Severe Aphasia, 12th MMY
Children's Articulation Test, 12th MMY
CID Phonetic Inventory, 11th MMY
CID Picture SPINE, 11th MMY
Clinical Evaluation of Language Fundamentals, Third Edition, 13th MMY
Clinical Evaluation of Language Fundamentals—Preschool, 13th MMY
Communication Activities of Daily Living, Second Edition, 14th MMY
Communication and Symbolic Behavior Scales, 13th MMY
Communication Profile: A Functional Skills Survey, 14th MMY
Communicative Abilities in Daily Living, 10th MMY
Comprehensive Screening Tool for Determining Optimal Communication Mode, 10th MMY
Computer Managed Articulation Diagnosis, 11th MMY
Computer Managed Screening Test, 11th MMY
Cooper Assessment for Stuttering Syndromes, 14th MMY
Denver Audiometric Screening Test, 9th MMY
Draw A Person: A Quantitative Scoring System, 11th MMY
Dysarthria Examination Battery, 14th MMY
Dysarthria Profile, 11th MMY
Early Speech Perception Test, 12th MMY
Elicited Articulatory System Evaluation, 11th MMY
Examining for Aphasia, Third Edition, 13th MMY
Fluharty Preschool Speech and Language Screening Test, 9th MMY
Frenchay Aphasia Screening Test, 11th MMY
Frenchay Dysarthria Assessment, 12th MMY
Goldman-Fristoe Test of Articulation, 10th MMY
Hearing Measurement Scale (The), 9th MMY
Hundred Pictures Naming Test (The), 12th MMY
INteraction CHecklist for Augmentative Communication, Revised Edition, 13th MMY
Iowa's Severity Rating Scales for Speech and Language Impairments, 12th MMY
Joliet 3-Minute Preschool Speech and Language Screen, 14th MMY
Joliet 3-Minute Speech and Language Screen (Revised), 13th MMY

Jones-Mohr Listening Test (The), 9th MMY
Khan-Lewis Phonological Analysis, 10th MMY
Lindamood Auditory Conceptualization Test, Revised Edition, 9th MMY
Metaphon, 12th MMY
Multilingual Aphasia Examination, Third Edition, 14th MMY
Northwestern Syntax Screening Test, 9th MMY
Oral Motor Assessment and Treatment: Improving Syllable Production, 10th MMY
Oral Speech Mechanism Screening Examination, Third Edition, 14th MMY
Oral Speech Mechanism Screening Examination—Revised, 11th MMY
PACS Pictures: Language Elicitation Materials, 11th MMY
Patterned Elicitation Syntax Test with Morphonemic Analysis (The), 13th MMY
Phonological Assessment of Child Speech, 10th MMY
Phonological Process Analysis, 12th MMY
Phonological Processes Assessment and Intervention, 11th MMY
Photo Articulation Test, Third Edition, 14th MMY
Prosody-Voice Screening Profile, 12th MMY
Quick Screen of Phonology, 12th MMY
Receptive-Expressive Emergent Language Test, Second Edition, 12th MMY
Revised Edinburgh Functional Communication Profile, 12th MMY
Revised Evaluating Acquired Skills in Communication, 12th MMY
Roswell-Chall Auditory Blending Test, 14th MMY
Scales of Early Communication Skills for Hearing-Impaired Children, 12th MMY
SCAN: A Screening Test for Auditory Processing Disorders, 11th MMY
SCAN-A: A Test for Auditory Processing Disorders in Adolescents and Adults, 13th MMY
Screening Instrument for Targeting Educational Risk, 12th MMY
Screening Test for Developmental Apraxia of Speech, 11th MMY
Slosson Articulation Language Test with Phonology, 12th MMY
Smit-Hand Articulation and Phonology Evaluation, 14th MMY
Speech and Language Evaluation Scale, 12th MMY
STIM/CON: Prognostic Inventory for Misarticulating Kindergarten and First Grade Children, 11th MMY
Structured Photographic Articulation Test Featuring Dudsberry, 12th MMY
Stuttering Severity Instrument for Children and Adults, Third Edition, 13th MMY
Temple University Short Syntax Inventory, 12th MMY
Test for Auditory Comprehension of Language, Revised Edition, 10th MMY
Test for Examining Expressive Morphology, 10th MMY
Test of Articulation in Context, 14th MMY

Test of Auditory-Perceptual Skills, 13th MMY
Test of Oral Structures and Functions, 12th MMY
Test of Phonological Awareness, 13th MMY
Test of Pragmatic Language, 12th MMY
Voice Assessment Protocol for Children and Adults (A), 11th MMY
Wepman's Auditory Discrimination Test, Second Edition, 11th MMY
Wiig Criterion Referenced Inventory of Language, 13th MMY

7.14.19 Vocations

Access Management Survey, 12th MMY
Accounting Aptitude Test, 13th MMY
Accounting Program Admission Test, 12th MMY
Adult Career Concerns Inventory, 12th MMY
Allied Health Professions Admission Test, 12th MMY
American Occupational Therapy Association, Inc. Fieldwork Evaluation for the Occupation Therapist (The), 12th MMY
Applied Knowledge Test, 9th MMY
Aptitude Interest Inventory, 13th MMY
Armed Services-Civilian Vocational Interest Survey, 10th MMY
Ashland Interest Assessment, 14th MMY
ASPIRE (A Sales Potential Inventory for Real Estate), 12th MMY
ASSESS Personality Battery [Expert System Version 5. X], 14th MMY
Assessing Specific Competencies, 12th MMY
Assessing Your Team: Seven Measures of Team Success, 14th MMY
Assessment Inventory for Management, 13th MMY
Assessment of Career Decision Making, 10th MMY
Assessment of Competencies for Instructor Development, 11th MMY
Automated Office Battery, 11th MMY
Ball Aptitude Battery, 9th MMY
Becker Work Adjustment Profile, 11th MMY
Belbin Team-Roles Self-Perception Inventory, 14th MMY
BENCHMARKS, 12th MMY
Bennett Mechanical Comprehension Test, 11th MMY
BRIGANCE Diagnostic Employability Skills Inventory, 14th MMY
Business Analyst Skills Evaluation, 10th MMY
Campbell Interest and Skill Survey, 13th MMY
Campbell Organizational Survey, 12th MMY
Campbell-Hallam Team Development Survey, 14th MMY

Canadian Occupational Interest Inventory, 9th MMY
Candidate Profile Record (The), 12th MMY
Career Anchors: Discovering Your Real Values, Revised Edition, 13th MMY
Career Assessment Inventory, Second Edition (Vocational Version), 11th MMY
Career Assessment Inventory—The Enhanced Version, 10th MMY
Career Attitudes and Strategies Inventory: An Inventory for Understanding Adult Careers, 13th MMY
Career Decision Scale, 9th MMY
Career Decision-Making Self-Efficacy Scale, 14th MMY
Career Development Inventory [Consulting Psychologists Press, Inc.], 13th MMY
Career Directions Inventory, 10th MMY
Career Exploration Inventory (The), 13th MMY
Career Exploration Series, 1992 Revision, 13th MMY
Career Factors Inventory, 14th MMY
Career Guidance Inventory, 11th MMY
Career Interest Inventory, 13th MMY
Career Interests Test, 9th MMY
Career IQ Test, 14th MMY
Career Occupational Preference System—Professional Level Interest Inventory, 14th MMY
Career Problem Check List, 9th MMY
Career Profile System, Second Edition (The), 13th MMY
Career Thoughts Inventory, 14th MMY
Career Values Card Sort, 13th MMY
Change Abilitator, 14th MMY
Chronicle Career Quest, 12th MMY
Clerical Abilities Battery, 11th MMY
Clerical Staff Selector, 10th MMY
Coaching Process Questionnaire, 13th MMY
Common-Metric Questionnaire (The), 13th MMY
Communication Knowledge Inventory, 11th MMY
Communications Profile Questionnaire, 10th MMY
Competency-Based Performance Improvement: Organizational Assessment Package, 14th MMY
Computer Programmer Aptitude Battery, 11th MMY
Computer Programmer Test Package, 11th MMY
Cosmetology Student Admissions Evaluation, 11th MMY
Dental Admission Test, 12th MMY

Developing the High-Performance Workplace, 14th MMY
Developmental Challenge Profile: Learning from Job Experiences, 13th MMY
Devine Inventory (The), 12th MMY
Diversity Management Survey (The), 14th MMY
Electrical Maintenance Trainee, 13th MMY
Electromechanical Vocational Assessment Manuals, 13th MMY
Electronic and Instrumentation Technician Test, 10th MMY
Electronics Test—Form G2, 13th MMY
Employability Inventory (The), 12th MMY
Employability Maturity Interview, 11th MMY
Employee Aptitude Survey, Second Edition, 14th MMY
Employee Effectiveness Profile, 12th MMY
Employment Screening Test and Standardization Manual, 11th MMY
Employment Values Inventory, 14th MMY
Empowerment Inventory, 13th MMY
Endicott Work Productivity Scale, 14th MMY
Entrance Examination for Schools of Nursing [RN Entrance Examination], 12th MMY
Entrance Examination for Schools of Practical/Vocational Nursing, 12th MMY
EQ Questionnaire, 14th MMY
ETSA Tests, 11th MMY
Evaluating the Participant's Employability Skills, 12th MMY
Filipino Work Values Scale (The), 14th MMY
Firefighter Selection Test, 11th MMY
Fleishman Job Analysis Survey, 12th MMY
General Clerical Test, 12th MMY
General Management In-Basket, 12th MMY
Gordon Occupational Check List II, 9th MMY
Guilford-Zimmerman Interest Inventory (The), 12th MMY
Hall Occupational Orientation Inventory, 12th MMY
Harrington-O'Shea Career Decision-Making System Revised (The), 12th MMY
Hay Aptitude Test Battery, 12th MMY
Hay Aptitude Test Battery [Revised], 14th MMY
High School Career-Course Planner, 12th MMY
Hilson Personnel Profile/Success Quotient, 11th MMY
Hogan Personnel Selection Series, 11th MMY
IDEAS: Interest Determination, Exploration and Assessment System, 11th MMY
Incentives Management Index, 14th MMY
Instrument Technician, 13th MMY

Interest Determination, Exploration and Assessment System. *See* IDEAS: Interest Determination, Exploration and Assessment System

Intuitive Mechanics (Weights & Pulleys), 9th MMY

Inventory of Interests, 9th MMY

IPI Job-Tests Program, 11th MMY

IPMA Fire Service Tests, 9th MMY

Jackson Vocational Interest Survey [1995 Revision], 14th MMY

JIIG-CAL Occupational Interests Guide, 9th MMY

Job Activity Preference Questionnaire, 9th MMY

Job Awareness Inventory, 9th MMY

Job Choice Decision-Making Exercise (A), 11th MMY

Job Descriptive Index and Retirement Descriptive Index, 12th MMY

Job Search Inventory, 13th MMY

Job Seeking Skills Assessment, 12th MMY

Job Style Indicator, 13th MMY

Job Training Assessment Program, 10th MMY

Jobmatch, 9th MMY

JOB-O, 12th MMY

Keyboard Skills Test, 11th MMY

KEYS: Assessing the Climate for Creativity, 14th MMY

Kirkpatrick Management and Supervisory Skills Series (The), 14th MMY

Kuder General Interest Survey, Form E, 12th MMY

Kuder Occupational Interest Survey, Revised (Form DD), 1985, 10th MMY

Law Enforcement Assessment and Development Report, 10th MMY

Law Enforcement Candidate Record, 13th MMY

Law School Admission Test, 13th MMY

Leader Behavior Analysis II, 12th MMY

Leader Behavior Questionnaire, Revised, 14th MMY

Leadership Practices Inventory—Delta, 14th MMY

Leadership Skills Inventory [Consulting Resource Group], 13th MMY

Learning Organization Practices Profile, 14th MMY

Leatherman Leadership Questionnaire [Revised], 12th MMY

Leatherman Leadership Questionnaire: 360 Degree Leadership Profile, 14th MMY

Leisure Search Inventory, 13th MMY

Life Style Questionnaire, 9th MMY

Lore Leadership Assessment, 14th MMY

Maintest, 13th MMY

Major-Minor-Finder, 1986-1996 Edition (The), 10th MMY

Management and Graduate Item Bank, 11th MMY
Management & Leadership Systems, 14th MMY
Management Appraisal Survey, 10th MMY
Management Change Inventory, 11th MMY
Management Interest Inventory, 11th MMY
Management Inventory on Leadership, Motivation and Decision-Making, 12th MMY
Management Inventory on Modern Management, 10th MMY
Management Style Inventory [Hanson Silver Strong & Associates], 13th MMY
Management Styles Inventory, 12th MMY
Management Styles Questionnaire, 10th MMY
Manager Profile Record, 11th MMY
Manager Style Appraisal, 12th MMY
Manager/Supervisor Staff Selector, 10th MMY
Managerial and Professional Job Functions Inventory, 11th MMY
Managerial Assessment of Proficiency MAP, 11th MMY
Managerial Competence Index, 11th MMY
Managerial Style Questionnaire, 10th MMY
MAPP: Motivational Appraisal of Personal Potential, 14th MMY
MbM Questionnaire: Managing by Motivation, Third Edition (The), 14th MMY
MDS Vocational Interest Exploration System, 13th MMY
Measurement of Counselor Competencies: A Self-Assessment, 10th MMY
MecTest (A Test for Maintenance Mechanics), 13th MMY
Microcomputer User Aptitude Test, 11th MMY
Minnesota Clerical Assessment Battery, 14th MMY
Minnesota Clerical Test, 9th MMY
Minnesota Importance Questionnaire, 11th MMY
Minnesota Spatial Relations Test, Revised Edition, 9th MMY
Modern Occupational Skills Tests, 12th MMY
Motivated Skills Card Sort, 13th MMY
Motivational Appraisal of Personal Potential. *See* MAPP: Motivational Appraisal of Personal Potential
Multifactor Leadership Questionnaire (Revised), 12th MMY
Multifactor Leadership Questionnaire for Research, 14th MMY
My Vocational Situation, 9th MMY
National Police Officer Selection Test, 12th MMY
New Technology Tests: Computer Commands, 11th MMY
New Technology Tests: Computer Rules, 11th MMY
NOCTI Teacher Occupational Competency Test: Appliance Repair, 13th MMY

NOCTI Teacher Occupational Competency Test: Audio-Visual Communications Technology, 13th MMY
NOCTI Teacher Occupational Competency Test: Auto Body Repair, Forms B and D, 13th MMY
NOCTI Teacher Occupational Competency Test: Automotive Technician, 13th MMY
NOCTI Teacher Occupational Competency Test: Brick Masonry, 10th MMY
NOCTI Teacher Occupational Competency Test: Building and Home Maintenance Services, 10th MMY
NOCTI Teacher Occupational Competency Test: Child Care and Guidance, 12th MMY
NOCTI Teacher Occupational Competency Test: Machine Trades, 10th MMY
NOCTI Teacher Occupational Competency Test: Mechanical Technology, 12th MMY
NOCTI Teacher Occupational Competency Test: Microcomputer Repair, 12th MMY
NOCTI Teacher Occupational Competency Test: Scientific Data Processing, 13th MMY
Nurse Aide Practice Test, 12th MMY
Occupational Aptitude Survey and Interest Schedule, Second Edition—Aptitude Survey, 12th MMY
Occupational Aptitude Survey and Interest Schedule, Second Edition—Interest Schedule, 12th MMY
Occupational Interests Card Sort, 13th MMY
Occupational Interests Explorer (The), 12th MMY
Occupational Interests Surveyor (The), 12th MMY
Occupational Personality Assessment, 13th MMY
Occupational Relationships Profile, 13th MMY
Occupational Test Series—Basic Skills Tests, 12th MMY
Occupational Test Series: General Ability Tests, 11th MMY
Office Skills Series, 13th MMY
Optometry Admission Testing Program, 12th MMY
Organizational Beliefs Questionnaire, 14th MMY
Organizational Change-Readiness Scale, 14th MMY
Orientation Test for Professional Accounting, 12th MMY
PAR: Proficiency Assessment Report, 12th MMY
Parker Team Player Survey, 13th MMY
PDI Employment Inventory and PDI Customer Service Inventory, 14th MMY
People Performance Profile, 10th MMY
Performance Skills Leader, 14th MMY
Personal Opinion Matrix, 10th MMY
Personnel Assessment Selection System, 11th MMY
Personnel Reaction Blank, 13th MMY

Personnel Relations Survey [Revised], 14th MMY
Pharmacy College Admission Test, 13th MMY
Picha-Seron Career Analysis, 12th MMY
Police Selection Test, 14th MMY
Poppleton Allen Sales Aptitude Test (The), 10th MMY
Position Analysis Questionnaire, 12th MMY
Power Base Inventory, 13th MMY
Prevocational Assessment Screen, 12th MMY
Process Technician Test, 10th MMY
Production and Maintenance Technician Test, 11th MMY
Productivity Assessment Questionnaire, 10th MMY
Professional and Managerial Position Questionnaire, 12th MMY
Professional Employment Test, 12th MMY
Proficiency Assessment Report. *See* PAR: Proficiency Assessment Report
Profile of Aptitude for Leadership, 13th MMY
Programmer Analyst Aptitude Test, 11th MMY
Project Management Knowledge & Skills Assessment, 14th MMY
Prospector: Discovering the Ability to Learn and to Lead, 14th MMY
PSC Survey-SA, 13th MMY
PSC-Survey ADT, 13th MMY
Rahim Organizational Conflict Inventories, 10th MMY
Ramsay Corporation Job Skills—Machinist Test, 13th MMY
Ramsay Corporation Job Skills—Mechanic Evaluation Test, 13th MMY
Ramsay Corporation Job Skills—Reading Electrical Drawings & Schematics, 13th MMY
Ramsay Corporation Job Skills—Reading Prints & Drawings, 13th MMY
Reading-Free Vocational Interest Inventory, Revised, 11th MMY
Reality Check Survey, 12th MMY
Revised Minnesota Paper Form Board Test, Second Edition, 14th MMY
Revised PSB—Aptitude for Practical Nursing Examination, 13th MMY
Revised PSB—Health Occupations Aptitude Examination, 13th MMY
Revised PSB—Nursing School Aptitude Examination (R. N.), 13th MMY
Roeder Manipulative Aptitude Test, 12th MMY
Rothwell Miller Interest Blank (Revised), 13th MMY
Rothwell Miller Values Blank, 14th MMY
Safran Student's Interest Inventory, Third Edition, 12th MMY
Sales Achievement Predictor, 14th MMY
Sales Motivation Inventory, Revised, 11th MMY

Sales Motivation Survey, 14th MMY

Sales Potential Inventory for Real Estate. *See* ASPIRE (A Sales Potential Inventory for Real Estate)

Sales Staff Selector, 10th MMY

Sales Style Indicator, 13th MMY

Selby MillSmith Values Indices, 14th MMY

SELECT Associate Screening System, 14th MMY

Self-Directed Search, 4th Edition [Forms R, E, and CP], 13th MMY, 14th MMY

Self-Motivated Career Planning, 11th MMY

ServiceFirst, 14th MMY

Short Employment Tests, Second Edition, 13th MMY

Situation Diagnosis Questionnaire, 10th MMY

Skills Assessment Module, 11th MMY

Skills Inventory for Teams, 13th MMY

SkillScan for Management Development, 14th MMY

SKILLSCOPE for Managers, 12th MMY

Standard Timing Model (The), 12th MMY

Stenographic Skill-Dictation Test, 12th MMY

Strong Interest Inventory [Fourth Edition], 12th MMY

Styles of Leadership Survey, 12th MMY

Styles of Management Inventory, 12th MMY

Supervisory Communication Relations, 11th MMY

Supervisory Profile Record, 11th MMY

Survey of Employee Access, 12th MMY

Survey of Work Styles, 13th MMY

Systems Programming Aptitude Test, 10th MMY

Team Effectiveness Survey, 12th MMY

Team Leadership Practices Inventory (The), 13th MMY

Team Process Diagnostic, 11th MMY

Teamness Index, 11th MMY

Technical Test Battery, 11th MMY

Technician Electrical Test—Form A2, 13th MMY

Technician Mechanical Test—Form A2, 13th MMY

Time of Your Life (The), 12th MMY

Tool Room Attendant, 13th MMY

Total Quality Management Readiness Index, 13th MMY

Trainer's Assessment of Proficiency (TAP), 12th MMY

Transit Operator Selection Inventory, 11th MMY

Transition-to-Work Inventory: A Job Placement System for Workers with Severe disabilities, 14th MMY
Typing Skill Test, 11th MMY
USES Basic Occupational Literacy Test, 11th MMY
Veterinary College Admission Test, 12th MMY
Vocational Assessment and Curriculum Guide, 13th MMY
Vocational Decision-Making Interview—Revised, 14th MMY
Vocational Integration Index, 14th MMY
Vocational Interest, Experience and Skill Assessment (VIESA), Canadian Edition, 12th MMY
Vocational Interest Inventory—Revised, 13th MMY
Vocational Interest Inventory and Exploration Survey, 12th MMY
Vocational Preference Inventory, 1985 Edition, 10th MMY
Vocational Research Interest Inventory, 11th MMY
Voc-Tech Quick Screener, 12th MMY
Walden Structured Analysis and Design Concepts Proficiency Test, 11th MMY
Welding Test, 13th MMY
Word Processor Assessment Battery, 10th MMY
Work Adjustment Inventory, 13th MMY
Work Adjustment Scale (The), 13th MMY
Work Aptitude: Profile and Practice Set, 10th MMY
Work Experience Survey (The), 14th MMY
Work Performance Assessment, 13th MMY
Work Skills Series Production, 12th MMY
Work Temperament Inventory, 13th MMY
Worker Rehabilitation Questionnaire, 14th MMY
Working—Assessing Skills, Habits, and Style, 14th MMY

7.15 研究者如何创建自己的量表:最后一项工作

研究者创建新量表的可能情况有两种:一是,在做了大量文献研究后,没有发现适合自己所研究问题的量表;二是,可利用的量表结构不合理。任何一份结构式量表应该有助于日后使用该量表的研究者,在调查过程中得到更多的研究发现。将一些项目简单合并和加以权重处理并不是创建量表。一份像样的量表必须有严谨的理论结构,并通过既定评估标准的研讨。

一些研究者仅根据自己的研究旨趣而建构一些量表。他们并没有充分重视量表

在心理测验中的信度和效度。这样的量表用来检验研究假设时,由于在调查问题上很难做到明确无偏向的判断,往往得出无效的结果。

下文是从 Robert Athanasiou 和 Kendra B. Head 建构的系列量表中选择的项目(Robinson, Athanasious, & Head, 1969, pp. 4-13)。

Ⅰ. 项目建构准则

1. 选择的项目能够准确全面地反映被测变量所包含项目的全体。
2. 陈述项目的用语浅显易懂,让受访者能够比较容易理解、填答量表。
3. 项目分析显示,每个项目都与所测量的变量紧密相关。项目分析的技术包括如下几点:
   - 项目相关矩阵
   - 因子分析
   - 项目相关的客观标准
4. 对不符合要求的项目进行预测试、淘汰或修正。

Ⅱ. 答案设置标准

1. 为了避免由于经验知识(顺从的毛病)设置有倾向性的备选项,可使用下面提到的一些设计策略:
   - 淘汰简单讨论通过的项目。
   - 偶尔改变在肯定和否定之间两者选一的答案。
   - 使用单项选择题:给出两个或更多的备选项,让回答者只从中选择一项。
2. 为了避免因迎合社会需要(好印象的毛病)而设置有偏向的选项,可以使用下面提到的一些设计方法:
   - 运用迫选项目,这种项目的各备选项在迎合社会需要方面不存在明显差别。
   - 对表示社会愿望的选项进行预先测试,不能证明其与社会愿望相一致的选项予以舍弃、修正。
3. 分析回答者的答案,对那些回答者根据自己意想编造的问卷给予作废处理。
4. 分析回答者的答案,对那些由于回答者为伪造出答案具有前后一致性,而选择少数几种备选项或一些极端值备选项的问卷,给予作废处理。

Ⅲ. 量表测量标准

1. 典型抽样:抽样方法的恰当,可以使样本类似于一个缩微的总体,并能够全面反映要研究的总体特征。
2. 充足的正规信息:为了保证被调查者在接受调查时,能够准确地理解量表,研究者应该考虑以下结构式问卷的测量。
   - 确立样本所依据的量表均值和标准差。
   - 明确界定的选项组的均值和标准差。
   - 项目的均值和标准差。
3. 信度:一份可靠的量表可以对其测量的对象保持测量结果的一致性。信度的确定可以参考下面三条标准。

- 被调查者在两次不同的时间回答同一项目所得分值具有高度相关性。(再测信度)
- 利用两组不同的项目同时测量调查对象,测量结果的相关系数很高(这些项目可以用两份一致性的量表呈现出来,也可以在同一份量表中分成两组呈现)。(折半信度)
- 所有被调查者回答这两份量表项目的结果相关系数很高。(复本信度)

4. 同质性:一份同质性量表具有一致性。同质性是指量表内部的连贯性。这一特性对量表结构非常重要,只有同质性量表才可以表现出一般性特征。测量量表的同质性包括下述几种方法:
   - 折半相关性。
   - 内部一致性量表,形式相似、项目指标相关。
   - Guttuman 量表的方法。
5. 效度:一份有效的量表可以对其测量对象有效测量。效度(最好能够就相关系数做外部标准的测量)意味着量表提供的预测能力超越了量表构成要素直接涉及的范围。量表效度的测量方法有如下几种:
   - 已知组之间的差异。
   - 双重交互认定。
   - 多重特性,多峰矩阵。
   - 对被测量的现象有一个或多个标准的量表,其分值之间的相关性。

译者后记

对于从事社会科学经验研究来说,研究设计和变量测量无疑是两个相当关键的环节。一项具体的经验研究如同一项工程,事前的设计是研究者理论素养、方法素养、逻辑思维、社会知识以及研究艺术的综合体现。在一定意义上,我们甚至可以说,研究设计决定了研究的质量、研究的收益,甚至研究的成败。至于社会测量,则是社会科学研究的"瓶颈"。与自然现象那种相对具体、相对客观、相对单纯不同,社会现象显然更加抽象、更加主观,也更加复杂,对于要探索其奥秘的社会研究者来说,如何更有效地对社会现象进行测量,一直是困扰他们的首要问题。

本书正是抓住这两个重要的方面,为社会研究者学习和查阅相关的方法技术提供了很好的帮助。该书自上世纪 70 年代初出版以来,深受广大社会研究者欢迎,到 2002 年已出版了第 6 版。将此书翻译出来,对我国社会学以及政治学、人口学、教育学、新闻传播学、管理学、心理学、社会工作等专业领域的经验研究将会起到一定的促进作用。特别是书中大量的量表和指数,十分便于研究者借鉴和使用。

本书各部分的译者是:

风笑天(社会学博士,南京大学社会学系教授,译序言、致谢、第 1 章、第 2 章)

邹宇春(社会学硕士,广州市社会科学院社会学所助研,译第 3 章第 1 ~7 节)

邓希泉(社会学硕士,中国青少年研究中心助研,译第 3 章第 8 节)

史伟伟(香港科技大学社科部研究生,译第 4 章)

吴元清(社会学硕士,杭州师范学院社会工作系讲师,译第 5 章第 1 ~6 节)

李良进(社会学硕士,深圳职业技术学院教师,译第 5 章第 7 ~11 节)

蔡扬眉(社会学硕士,北京工业大学社会工作系助教,译第 6 章)

张静波(南京大学社会学系硕士研究生,译第 7 章第 1、2 节及第 3 节第 1 目)

杜荟苑(南京大学社会学系硕士研究生,译第 7 章第 3 节第 2 目至第 4 节第 3 目)

江　臻(南京大学社会学系硕士研究生,译第 7 章第 4 节第 4 目至第 8 节)

童宗斌(南京大学社会学系硕士研究生,译第 7 章第 9 ~15 节)

译稿完成后,由我和张静波对译稿进行了校订,我的研究生杜林、汤华、王梁也协助我作了部分资料的初译工作。译文中的不当之处,欢迎读者批评指正。

风笑天

2004 年 7 月 31 日于南京